

THESIS

A KINETIC MODEL DEVELOPMENT OF THE M13 BACTERIOPHAGE LIFE CYCLE

Submitted by

Steven William Smeal

Department of Chemical and Biological Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2014

Master's Committee:

Advisor: Nick Fisk

Co-Advisor: Ashok Prasad

Claudia Gentry-Weeks

Copyright by Steven William Smeal 2014

All Rights Reserved

ABSTRACT

A KINETIC MODEL DEVELOPMENT OF THE M13 BACTERIOPHAGE LIFE CYCLE

A kinetic model which can simulate the M13 bacteriophage (a virus which only infects bacteria) life-cycle was created through a set of ordinary differential equations. The M13 bacteriophage is a filamentous phage with a circular single-stranded DNA genome. The kinetic model was developed by converting the biology into ordinary differential equations through careful studying of the existing literature describing the M13 life cycle. Most of the differential equations follow simple mass-action kinetics but some have an additional function, called the Hill Function, to account for special scenarios. Whenever possible, the rate constants associated with each ordinary differential equation were based off of experimentally determined constants. The literature describing M13 viral infection did not provide all of the rate constants necessary for our model. The parameters which were not experimentally determined through literature were estimated in the model based on what is known about the process. At present, no experiments were performed by our lab to verify the model or expand on the information available in the literature. However, the M13 phage model has improved the understanding of phage biology and makes some suggestions about the unknown factors that are most important to quantitatively understanding phage biology.

The kinetic model is genetically structured and simulates all well-known and major features of viral phage infection beginning when the first viral ssDNA has entered the cytoplasm and ends right before the cell is ready to divide. The model includes DNA replication, transcription, translation, mRNA processing and degradation, viral protein P2 and viral DNA interaction, viral protein P5 and viral single-strand DNA (ssDNA) interaction, P5 and mRNA

interactions, and the assembly of new phage. Additionally, the model has implemented an interaction of P2 and P10, which has not been directly verified through experiments, to account for the negative effect P10 has on DNA replication. The interaction of the host cell and virus infection was not explicitly modeled, but a subset of cellular resources were set aside for phage reproduction based on experimental estimates of the metabolic burden of phage infection. Specifically, limited amount of host resources RNA polymerase, DNA polymerase 3, and ribosomes were allocated to phage reproduction. All other host resources such as nucleotides and amino acids were assumed to be in abundance and did not limit phage replication.

The model was verified by comparing the output of the model to a set of existing experimental results in literature. The model reproduced both the experimentally measured levels of phage proteins and mRNA, and the timing and dynamics of virus production for the first cell cycle after infection. All of the unknown parameters were based off the model results at the end of the first cell cycle. When the model was extended to account for phage production through multiple cell divisions, the model predicts the cell has the ability to cure itself from the infection in 7 – 8 cell cycles, which we found literature supporting our results after we made the conclusions.

Once the model was created we studied how host resources, RNA polymerase and ribosomes, were distributed during the infection process. We were also able to replicate an experiment describing the effects that the viral DNA binding protein P5 had on the translation of five other viral proteins in-silico. The role of P5 inhibition in the viral life-cycle is unclear and our in-depth analysis of P5 function has revealed a possible explanation of how P5 translational inhibition could be an evolutionary advantage. Additionally, we proposed a mechanism which has not been strongly suggested to exist in literature. We are anticipating the development of the

model will aid in the progress of phage display on filamentous phage and we believe the current model can be easily amendable to account for other phage like phages such as *Ike* filamentous phage. We discuss further additions and modifications to the model that will allow more exact treatment of early events in the phage life-cycle and more explicit coupling of phage life-cycle and host biology.

ACKNOWLEDGEMENTS

I have many people to thank for their roles in helping me to finish my project: Dr. Nick Fisk for guiding me through the research process and editing my thesis; Dr. Ashok Prasad who met with me on several occasions to discuss modeling; and all of the Fisk Lab members for helping me along the way by listening to each group presentation and giving valuable insights and questions about the model development. I am also especially appreciative and thankful of my parents, William E. Smeal and Kathleen Steffish, who both encouraged me to pursue a graduate degree and provided me with the support to accomplish it. Finally, I would like to thank my long-time girlfriend, Alyssa Macomber, who never complained when I worked late and was always there to listen.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	v
Chapter 1: Introduction.....	1
1.1 General Background on Viruses and Phage Biology.....	1
1.2 M13 Phage Structure and Genome.....	3
1.3 M13 Viral infection.....	6
1.3.1 Phage Absorption.....	6
1.3.2 DNA Replication.....	7
1.3.3 Transcription and mRNA Degradation.....	12
1.3.4 Translation.....	13
1.3.5 Phage Assembly.....	14
1.4 Biotechnological Applications of Bacteriophages.....	17
1.5 Why Model Biology.....	19
1.6 Examples of Previous Virus Life Cycles.....	20
1.7 Motivation Behind Building a Kinetic Model of M13.....	21
Chapter 2: Model Description.....	23
2.1 Summary.....	23
2.2 Formation of Viral dsDNA.....	24
2.3 P2 Nicking RF1 DNA.....	27
2.4 Production of ssDNA.....	29
2.5 P10 Inhibits DNA Synthesis and Phage Replication.....	31
2.6 Hill Equation Overview.....	32
2.7 Shift from RF1 to P5DNA Production.....	34
2.8 Transcription.....	37
2.9 RNAP Binding to the Promoter Site.....	40
2.10 RNAP Clearing the Promoter Site.....	42
2.11 Transcription Termination.....	43

2.12 mRNA Processing and Degradation.....	46
2.13 Translation.....	50
2.14 Rate of Ribosome Binding to Start Codon.....	50
2.15 Ribosomes Clearing the Promoter Site	54
2.16 Translation Termination.....	55
2.17 Assembly of New Phage	56
Chapter 3: Results and Discussion.....	61
3.1 Validation of the Model	61
3.1.1 Production of M13 Phage DNA	62
3.1.2 Production of mRNA Species.....	66
3.1.3 Protein Production from mRNA.....	71
3.1.4 Production of the Viral Species Involved in Regulation of Phage Infection.....	72
3.1.5 Production of the Viral Coat Species.....	75
3.1.6 Production of Assembly Proteins	79
3.1.7 Production of P5DNA.....	80
3.1.8 Production of Phage.....	82
3.2 Utilization of Host Resources.....	84
3.2.1 RNA Polymerase Utilization	85
3.2.2 Ribosome Utilization	89
3.3 Exploring the Effects of Removing P5 Inhibition of Proteins 1, 2, 3, 5, 10	93
3.4 Additional Control Over dsDNA Synthesis	101
3.4.1 Background Information.....	101
3.4.2 P10 Binding to ssDNA Mechanism.....	105
3.4.3 P2 Protecting the (-) Strand Origin of Replication Mechanism.....	107
3.4.4 Future Experiments.....	107
3.5 Cell Replication.....	109
3.6 Overestimation of the Rate of Phage Production	113
3.7 Expanded Pathway for mRNA Degradation	115
3.8 Possible RNAP Interference.....	117
Chapter 4: Conclusion.....	118

References.....	121
Appendix A – Simbiology Inputs	127
Appendix B – Simulation Strategy	133
Introduction to the Tools Used to Develop the Model.....	133
Files Needed for the Analysis of M13 Infection	134
General Strategy in the Creation of the m-Files.....	137
Appendix C: Source Code	143
ModelRun.m.....	143
Load_Parameters.m.....	146
Plot Phage Species.....	149
Appendix D: M13 Genome Sequence	154
LIST OF ABBREVIATIONS AND INITIAL PARAMTER VALUES.....	155

Chapter 1: Introduction

1.1 General Background on Viruses and Phage Biology

A virus contains genetic material protected by a protein coat. The genetic material of viruses contains the information necessary to reproduce the virus within a host. The genetic material of viruses can be in the form of single-stranded DNA (ssDNA), single-stranded RNA (ssRNA), double-stranded DNA (dsDNA), or double-stranded RNA (dsRNA) as opposed to cellular life that only uses dsDNA to store genetic information. Viruses do not have all of the required metabolic resources (e.g. nucleic acids, amino acids, etc.) or the cellular machinery (e.g. ribosomes, RNA polymerases, etc.), and must infect a host cell containing the required metabolic resources and cellular machinery in order to replicate. Therefore, viruses are obligate cellular predators or parasites. We are interested in a specific class of viruses called bacteriophage (phages for short). Specifically the thesis is about the M13 bacteriophage which infects *Escherichia coli* (*E. coli*).

Three outcomes typically occur when a bacteriophage genome enters a host cell. The host cell may recognize the phage genome as foreign nucleic acid and destroy it. The second outcome involves the viral genome being integrated within the host cell genome but not assembling new viral particles, which is called the lysogenic cycle. In the lysogenic cycle, the phage genome is passed on to daughter cells because it replicates with the host genome. Lastly the phage genome can initiate viral protein synthesis and the assembly of new viral particles, which the majority of phage infections results in the destruction of the host cell. When phage infection destroys the host cell the mode of phage replication is called lytic replication.

Any phage replicating through the lysogenic cycle can switch to the lytic cycle to assemble new phage. Therefore the lytic/lysogenic cycle can be considered one form of

replication. The lytic/lysogenic scenarios involve the phage completely hijacking all of the host resources to create phage progeny. The phage proteins are assembled inside the cell until the host cell bursts (called cell lysis) releasing all of the phage particles. Examples of lytic phages include the T4, lambda, and T7 viruses. Some temperate phage, such as the lambda phage, can reproduce in either the lytic or lysogenic cycle. Which cycle is chosen is a factor of environmental and host cell conditions.

The life cycle of the filamentous phage is unique among the known bacteriophages because it does not kill/lyse the host cell and infection. Filamentous phage replicates by co-existing with the cell such that infected cells continue to grow and divide while the phage particles are continuously produced. Assembly of the phage particle occurs at the cell membrane where the nascent phage is produced and extruded from the cell simultaneously. Since the phage particles are assembled at the cell membrane, no fully or partially assembled phage particle exists in the cells cytoplasm. The filamentous phage can co-exist with the cell because only a portion of the cellular resources are utilized by the phage. Cells infected with filamentous phage take longer to grow and divide than cells without the phage particle.

Phages come in many different shapes and can contain genetic information as single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), single-stranded RNA (ssRNA), or double-stranded RNA (dsRNA). The most common and diverse type of phage are in the taxonomic order of *Caudovirales* which includes the T4, lambda, and T7 phages. The *Caudovirales* all contain dsDNA, and the coat proteins forms an icosahedral head with a tail, see Figure 1 for a schematic of the *Caudovirales* structure. All known viruses within the order replicate through the lytic cycle [1].

Only four percent of known bacteriophages are not in the order of Caudovirales. These less common types of phages have several different morphologies [1]. The phage family *Tectiviridae*, example PRD1, contains a capsid head but no tail, dsDNA, and is lytic to the host cell. The *Tectiviridae* PRD1 contains an inner lipoprotein capsid. The phage family *Inoviridae*, examples f1, M13 and fd, are long and filamentous and can be rigid or flexible. They are released through excretion from the host cell and can co-exist for multiple generations. The phage family *Fuselloviridae*, example SSV1, has a shape which can take on multiple forms depending on the environmental conditions.

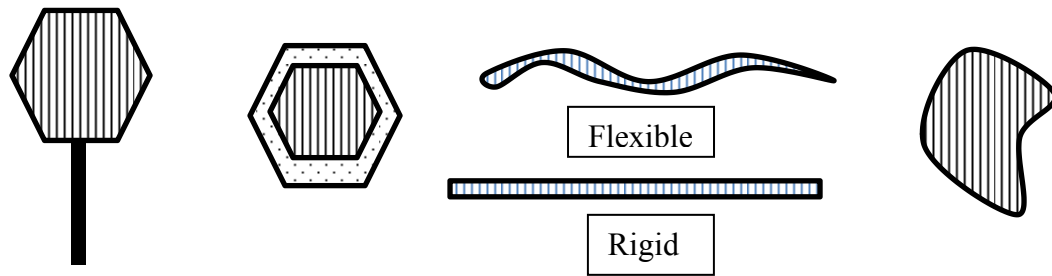


Figure 1 – examples of different bacteriophage structures, starting from the left include: *Caudovirales*, *Tectiviridae*, *Inoviridae* (Flexible or Rigid), and *Fuselloviridae*. Pictures were adapted from the following source [1].

The main focus of this work involves the development of a kinetic model of the life-cycle of the *Inovirus* M13. The M13 phage and two closely related phages, fd and f1, have genome sequence that are 98% homologous and the steps in viral replication are very similar [2]. The model described in the paper does not differentiate between M13, f1, and fd and can be considered as a basic model for these three phages.

1.2 M13 Phage Structure and Genome

M13 bacteriophage particle is a thin circular rod approximately 970 nm in length and 6-7 nm in diameter. It encapsulates a circular single-stranded DNA (ssDNA) genome which is 6407

nucleotides long and codes for 11 viral proteins. The M13 phage proteins are named numerically and will be designated in this thesis by P followed by a number. The viral coat consists of approximately 2700 copies of viral protein P8 (which is the major coat protein of the particle) and five copies of each of viral proteins P3, P6, P7, and P9, see Figure 2. The length of the particle and number of P8 molecules are directly related to the size of the genome packaged. Therefore, if modifications are done to the genome to increase the nucleotide length then the number of P8 molecules per phage particle will increase to ensure the entire genome is packaged. M13 viral infection does not kill the host cell, rather it can co-exists with the host cell through many generations and phage assembly occurs at the cellular membrane [3].

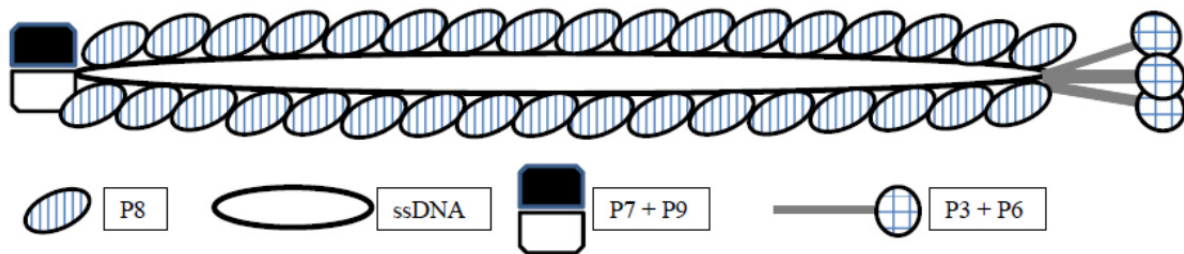


Figure 2 – A schematic diagram of the M13 bacteriophage structure.

The M13 viral genome is circular single-stranded DNA and can be separated into three functional units: the frequently transcribed region (FTR), lesser transcribed region (LTR), and the intergenic region (IR). The more frequently transcribed region contains the genetic information for viral proteins P2, P5, P7, P8, P9 and P10. Three mRNA transcripts are produced in the FTR and are initiated at the promoter sites labeled DA, DB, and DH as seen in Figure 3. All three transcripts produced from promoters DA, DB, and DH are terminated at the same

terminator site distal to P8. The terminator site distal to P8 is a strong rho-independent site and it is highly efficient at stopping transcription.

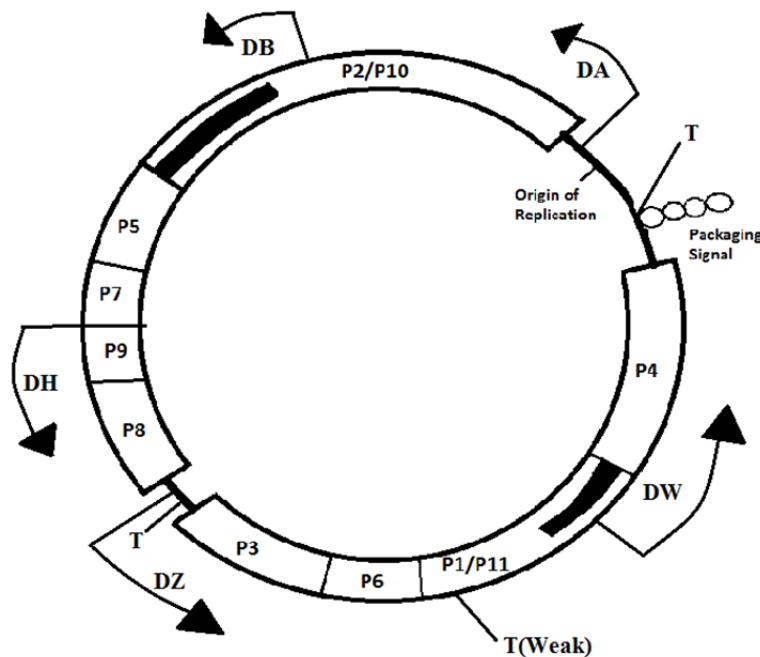


Figure 3 – A schematic of the viral genome where transcription proceeds in the counter-clockwise direction. It depicts the approximate locations of promoter sites DA, DB, DH, DZ, and DW and the three terminator sites. The picture is an adaptation from Russel, M. and Model P. 1988 [4].

The LTR contains two promoters, DW and DZ, which are transcribed less frequently than the FTR promoters. The promoter DZ can create two different mRNA transcripts which can code for only P3 and P6, or codes for P1, P3, P6, and P11. Figure 3 shows the promoter DZ is located between P8 and P3 coding regions. When transcription is initiated at the promoter site DZ it is terminated at the weak rho-dependent terminator site located distal to P6, which is successful 60 – 70 percent of the time [5]. When the terminator site does not work, transcription will continue and produce a transcript coding for two additional proteins P1 and P11 (the transcript codes for P3, P6, P1, and P11).

Looking at Figure 3, the next terminator following the one distal to P6 is located in the IR region and would indicate a possible transcript containing genes 3, 6, 1, 11, and 4. However, a transcript of this size has not been shown in-vivo and only the smaller transcript coding for genes 3, 6, 1, and 11 has been found [6]. Two possible explanation exists either there is an unknown terminator within P4 that has not been discovered or the transcript containing P4 is degraded very quickly. For the purposes of the model we are going to assume P4 is not part of the transcript coding for genes 3, 6, 1, and 11. P4 is transcribed by its own promoter site located upstream of P4 and is terminated in the intergenic region.

The intergenic region does not code for any proteins but it contains the origins of DNA replication and the packaging signal. The origin of replication is where DNA polymerase III (DP3) will bind to the genome to initiate DNA replication. There are two different regions for the initiation of DNA synthesis of the (-) and (+) strand of DNA. The packaging signal interacts with several of the phage encoded proteins in the cell membrane to initiate phage assembly.

1.3 M13 Viral infection

The M13 phage infection can be divided into five separate steps: phage absorption, DNA replication, transcription, translation, and phage assemble. Phage absorption involves a phage particle in the extracellular space inserting its ssDNA genome into an *E. coli* cell. DNA replication involves the formation of both dsDNA and ssDNA. Transcription and translation are required for viral protein synthesis. Finally, phage assembly will occur at the cellular membrane where new phage is assembled as they are extruded from the cell.

1.3.1 Phage Absorption

The M13 bacteriophage only infects *E. coli* containing the f-plasmid. Bacterial cells containing the f-plasmid are able to form a long hollow tube, called the f-pilus, which extrudes

from the cell membrane and can transfer genetic material to another cell. Phage infection is initiated when P3 binds to the end of the f-pilus, see Figure 4 . The f-pilus retracts to bring the viral particle closer to the host cell membrane. Once the virus reaches the membrane, several host proteins lead to passage of the viral genome into the cell and deposition of coat proteins in the inner membrane. P3 is also thought to recognize the Tol A protein located in the periplasmic space but the details of this process are not completely understood [7, 8]. The viral genome enters the host cell in a circularly closed single-stranded form.

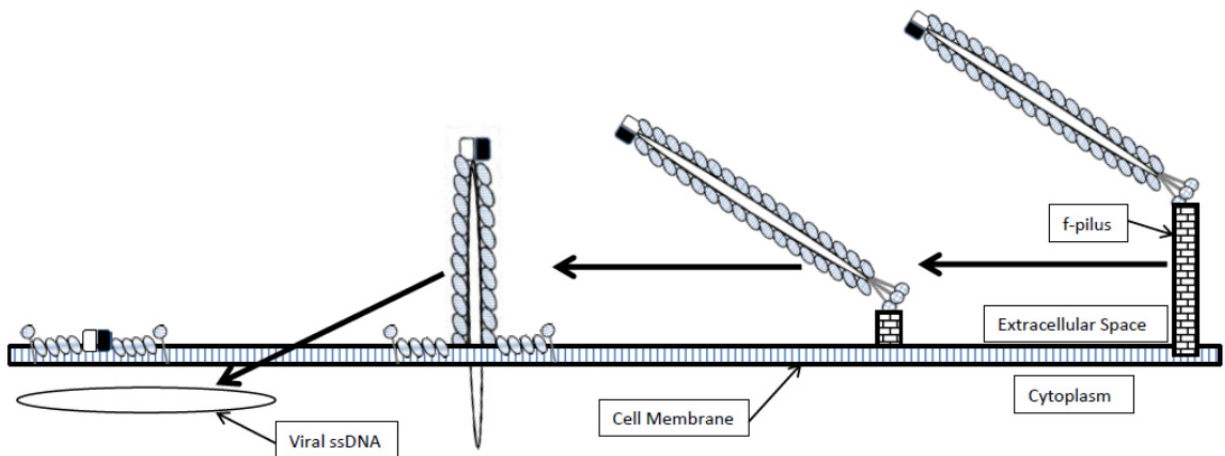


Figure 4 – Schematic of the steps for filamentous phage to insert the viral single-stranded DNA (ssDNA) into the host cells cytoplasm.

1.3.2 DNA Replication

The phage DNA replication process is complicated and involves several stages that are controlled by phage proteins. The first stage of DNA replication involves host enzymes (RNA polymerase (RNAP), DNA polymerase III (DP3), DNA polymerase I, and DNA ligase) converting the infecting ssDNA phage genome to a double-stranded DNA (dsDNA) form. The conversion of ssDNA to dsDNA is initiated by RNA polymerase which creates a short RNA primer on the ssDNA. Host DNA polymerase III (DP3) binds to the primed ssDNA and

synthesizes the complementary strand. DNA polymerase I removes the primer and fills the gap created by removing the primer. DNA ligase joins the two ends of the strands to create circular dsDNA.

Three forms of viral dsDNA are active in the host cell, supercoiled replicative form 1 (RF1), nicked and relaxed replicative form 2 (RF2), and relaxed and cyclized replicative form 4 (RF4) [9]. The different forms of dsDNA and the pathways through which they can interconvert can be seen in Figure 5. The product immediately following the conversion of the viral ssDNA to dsDNA is called RF4, which is relaxed DNA with no breaks in the (+) strand and no superhelical turns. Host enzyme DNA gyrase introduces superhelical turns into the DNA converting relaxed circular RF4 into supercoiled RF1. A break in the (+) strand of the supercoiled RF1 DNA is caused by the binding and catalytic action of two phage encoded P2 molecules. Nicking the plus strand (the strand with the same sequence as the entering genome) causes the dsDNA to unwind. The unwound DNA with a nick in the (+) strand and two molecules of P2 bound to the intergenic region is called RF2_i DNA. The P2 molecules on RF2_i DNA can either both dissociate to reform RF4 or only one P2 will dissociate to form RF2. RF2 is identical to RF2_i except that only one P2 molecule is bound to the (+) strand of the DNA. Host enzyme DNA polymerase 3 can then bind to RF2 to initiate DNA replication.

The initial ssDNA strand will be converted to RF1 DNA, which can then be used as a template for transcription but it cannot be used to initiate DNA replication. Phage DNA replication will not occur until viral protein P2 is synthesized and nicks RF1 DNA. The nick causes a break in the nucleic acid sequence on the (+) strand and relaxation of the supercoiled DNA. Once the nick has been created, DNA polymerase 3 (DP3), binds to the DNA and uses the (-) strand as a template for (+) strand synthesis. This mode of dsDNA replication via a single

stranded circular template is commonly called the rolling circle mechanism and is graphically shown in Figure 6.

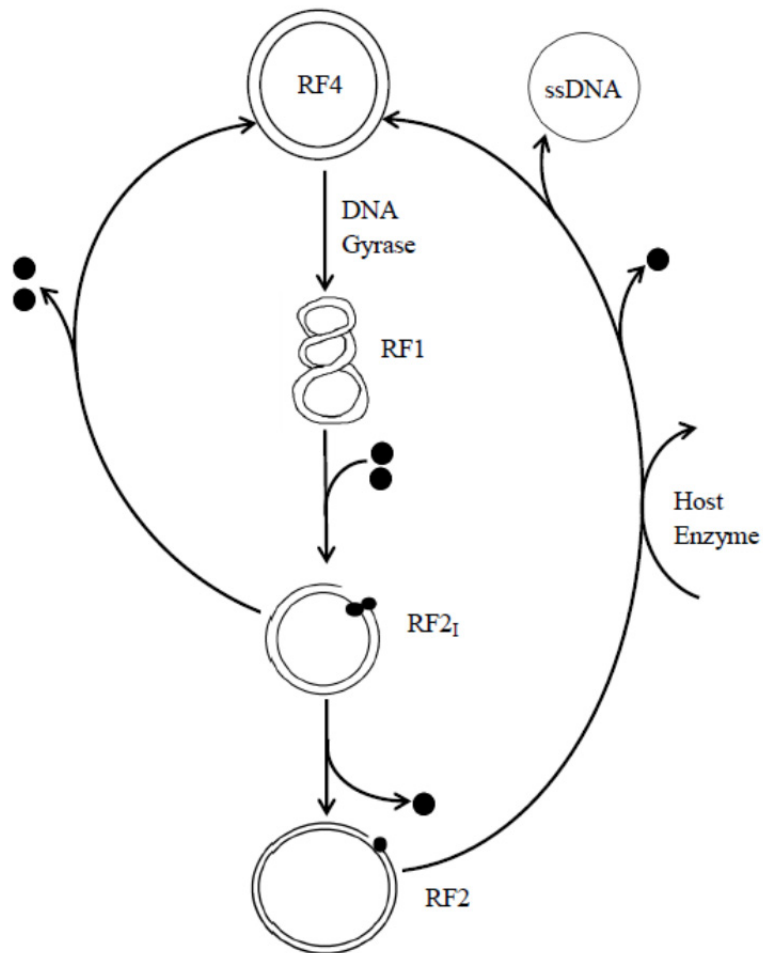


Figure 5– A schematic of how the conversions of the different forms of DNA interconvert between each other. It is a simplified version of M13 DNA pathway found in the following source [9]. The host enzymes include the RNA polymerase, DNA polymerase III (DP3), DNA polymerase I, and DNA ligase. ● - represents one P2 molecule.

The rolling circle mechanism is common in plasmids and several variations on the theme exist. Typically rolling circle replication involves an initiator protein binding to the super-coiled DNA and causing a break in one of the DNA strands. The initiator proteins have been observed to function as monomers or dimers. Although we previous discussed P2 acting as a dimer to

initiate DNA replication, it is actually unclear on whether it functions as a monomer or dimer. We presented it working as a dimer because the monomer mechanism is a small modification to Figure 6. If P2 nicks DNA as a dimer, the dimer will dissociate leaving a monomer bound to the DNA. In many rolling circle replication systems, once the DP3 enzymes initiates elongation, it may replicate the (+) strand several times before dissociating from the (-) strand. However, in M13 DNA replication DP3 is thought to only replicate one (+) strand before being dissociated from the DNA.

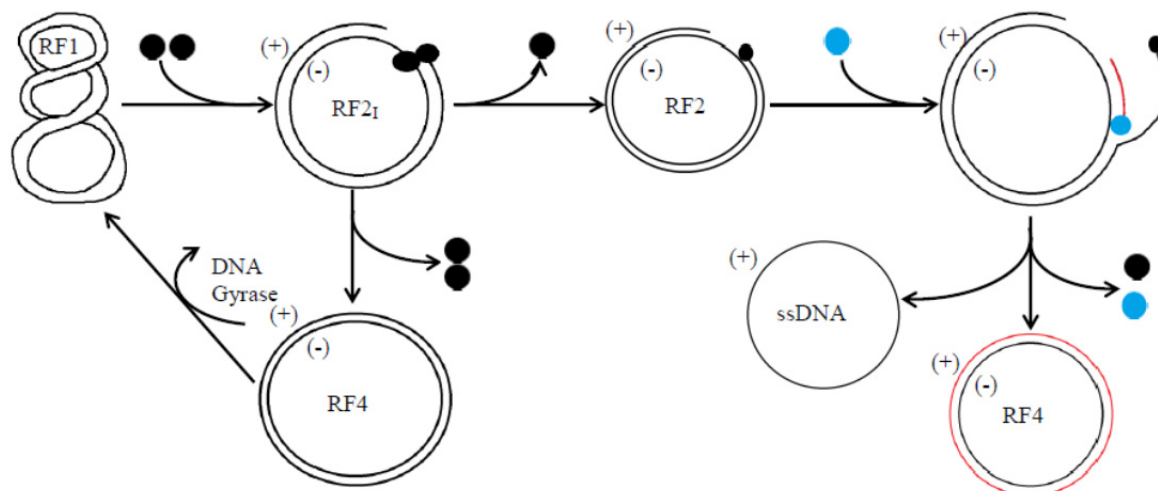


Figure 6 – A visual picture of the rolling circle DNA replication mechanism. ● - Represents one P2 molecule, ● - represents the DP3 and replicative helicase complex, and the red circle indicates the newly synthesized (+) strand.

A single round of replication produces a new ssDNA and converts the template from RF2 to RF4 dsDNA. The nascent RF4 DNA will be converted to RF1 by the successive action of DNA ligase and DNA Gyrase. The new RF1 DNA can be used as a template for transcription or DNA replication. The nascent ssDNA has two fates, it can either be converted to dsDNA by host enzymes or it can be sequestered by 1600 molecules of P5, we refer to the sequestered form of

ssDNA as P5DNA. The sequestering of ssDNA by P5 prevents the conversion to dsDNA. Experiments indicate that 30 – 200 strands of P5DNA accumulate by 30 minutes post infection [10-12]. P5DNA is the form of ssDNA that is packaged into newly formed phage particles. Early in the infection when levels of P5 are low, almost all ssDNA is converted to RF4 by host resources. As the number of molecules of P5 increases dsDNA synthesis decreases. Experiments involving DNA density labeling indicate that by 20 minutes post infection only 5% of phage DNA synthesis is devoted to dsDNA formation [13].

The viral protein P5 is important for the control of the accumulation of the dsDNA and for the assembly of phage particles. Two other proteins are involved in the accumulation of dsDNA and P5DNA formation, P2 and P10. In the absence of P2 or when the production of P10 is above wild-type levels, no DNA synthesis occurs. If P10 is not produced or production of P2 is above wild-type levels, no P5DNA is produced but dsDNA synthesis occurs even in the presence of the high amounts of the ssDNA binding protein P5 [14, 15]. These experimental observations indicate the combined actions of P2, P5, and P10 control the level of dsDNA and P5DNA synthesis.

In the absence of P2, DNA synthesis does not occur because RF1 DNA cannot be converted to RF2 DNA, an essential step in rolling circle replication. The molecular mechanism through which overproduction of P10 stops DNA synthesis is unclear and suggestions on the mechanism have been made [14, 15]. We have extended the ideas presented by Fulford and Model, 1988, and proposed a possible mechanism. Since the overproduction of P10 and the absence of P2 produce the same result then maybe P10 can inhibit P2 function by binding together to form an inactive complex P2P10. We are assuming one P2 molecule can bind to one P10 to form an inactive P2P10 binding complex. The assumption of P2 and P10 binding is

reasonable because P2 is thought to function as a dimer and P10 is translated from an in-frame start codon within the P2 gene. P10 is an essential protein for the accumulation of P5DNA and the role it plays is even more unclear, either this effect is directly related to the inhibition of P2 or involves other uncharacterized interactions. A possible mechanism for the P10 promotion of P5DNA formation is speculated upon in the results and discussion section.

1.3.3 Transcription and mRNA Degradation

Transcription of the dsDNA occurs at 5 promoter sites within the genome, and requires host enzyme RNA polymerase (RNAP). The three promoters DA, DB, and DH are all terminated at a single rho-independent termination site located distal to gene 8. DZ is terminated by a weak rho-dependent site distal to gene 6. Gene 4 has its own promoter, DW, and terminator site located in the intergenic region. Genes 1 and 11 lack a promoter to initiate transcription and must rely on the RNAP run-through of the rho-dependent terminator located distal to gene 6. The terminator is reported to be 60-70 percent efficient at stopping transcription [5]. DA, DB, and DH are considered strong promoters because transcription is initiated more frequently than promoters DZ and DW.

The level of transcript expression is not only dependent on transcript formation but also on degradation and the strength of ribosome binding. Degradation affects how many transcripts accumulate in the cell and ultimately protein synthesis. The degradation of mRNA species A, B, and H (created from promoter sites DA, DB, and DH respectively) have been extensively studied [16-24]. The mRNAs A and B have been shown to create smaller and more stable mRNAs [16-18, 22, 24, 25], table 1. The half-lives for mRNAs A-H have been determined in-vivo via labeling studies. Two different groups of enzymes can degrade mRNA, exonucleases degrade starting at the termini of the mRNA strand and endonucleases can cut at specific regions within

the mRNA transcript [26]. The degradation of mRNA species Z, Y, and W have not been as extensively studied [6, 27], and the half-lives are unknown. Therefore we assumed they will degrade in a single step.

Table 1 –The mRNA species A, B, H, Z, Y, and W are the transcripts directly transcribed from DNA. The other transcripts were degraded from the larger transcripts. Transcripts C-F all code for the same proteins but vary by transcript length and half-life [24]. Transcripts Z, Y, and W have not been extensively studied and no information is known about their degradation.

mRNA Name	Proteins Coded	Half-Life [Minutes]
A	P2, P10,P5,P7,P9 and P8	$\tau_A = < 2$
B	P10, P5,P7,P9 and P8	$\tau_B = < 2$
C	P5,P7,P9 and P8	$\tau_C = < 2$
D	P5,P7,P9 and P8	$\tau_D = < 2$
E	P5,P7,P9 and P8	$\tau_E = 2.5$
F	P5,P7,P9 and P8	$\tau_F = 6$
G	P9 and P8	$\tau_G = 8$
H	P9 and P8	$\tau_H = 10$
Z	P3 and P6	N/A
Y	P3, P6, P1, and P11	N/A
W	P4	N/A

1.3.4 Translation

The rate of protein synthesis is directly dependent on the level of production of the mRNA transcript and the rate at which ribosomes initiate at the start codon. The relative translation initiation frequencies (or ribosome binding site strengths/RBSS) have been estimated using the ribosome binding site calculator created by Salis lab (salis.psu.edu/software/). The input to the ribosome binding site calculator is the DNA sequence. From the DNA sequence, the calculator predicts the start codons of genes and returns a relative ribosome binding site strengths, which were estimated using a thermodynamic model [28, 29].

The ribosome initiation rate for proteins P1, P2, P3, P5, and P10 are all negatively inhibited by the single-stranded DNA binding protein, P5 [30, 31]. Inhibition is thought to occur

by P5 directly binding to the start codon which inhibits the ribosome to bind to it. P5 has been shown to inhibit the translation of P1, P2, P3, P5, and P10 with varies degrees of affect. Experiments have shown in the absence of P5 production, the translation of P2, P3, and P10 increase by a factor of 20.2, 18.3, and 18.7 respectively [30, 31]. In the same studies, P1 and P5 increase translation in the absence of P5 by a factor of 6.5 and 3.6 respectively.

Table 2 – The relative binding site strengths were calculated using the ribosome binding site calculator from Salis lab (salis.psu.edu/software/) [28, 29]. The input was the DNA sequence and the output was the ribosome binding site strengths relative to a promoter Salis lab chose. The relative biding site strength (RBBS) column converts the ribosome binding site strengths relative to the strongest promoter in the M13 genome, which is P5. The (*) indicates a RBSS which was not calculated through the ribosome binding site calculator. P10 was predicted to have a strong ribosome binding affinity by the calculator ($RBSS_{10} = 0.92$) and was suggested to have a strong binding site experimentally [16]. However, a secondary structure exists in the mRNA which inhibits the translation of P10 decreasing the efficiency of translation initiation 10-15 fold less than P2. Therefore the relative binding site strength for P10 was taken to be 10 fold less [19].

Protein	Nucleotide Length	Calculated RBSS	Relative Ribosome binding Site Strength (RBSS)	Protein Function
P2	1200	17078.77	0.54	Control
P10	350	29841.27	0.054 *	Control
P5	270	32146.26	1.00	Control
P7	101	0.50	1.55E-5	Coat
P9	98	2472.03	0.08	Coat
P8	220	6997.44	0.22	Coat
P3	1300	12828.44	0.40	Coat
P6	340	1422.46	0.04	Coat
P1	1040	2472.03	0.08	Assembly
P11	310	6337.81	0.20	Assembly
P4	1270	7372.44	0.23	Assembly

1.3.5 Phage Assembly

Phage assembly is the process of packaging ssDNA into phage particles which occurs at the cellular membrane for M13. Many of the details surrounding phage assembly are unclear because it is difficult to study events occurring in the membrane. What is known about the process are the proteins which are necessary for phage assembly to occur. A phage assembly site is a complex made up of phage proteins P1, P4, and P11 along with host protein thioredoxin. The

phage assembly complex spans both bacterial membranes. None of the assembly proteins are packaged with the phage particle. P4 will form a homomultimer complex in the shape of a cylindrical pore on the outer membrane, Figure 7. The homomultimer consists of about 14 molecules of P4 estimated from cryoelectron microscopic reconstruction [32]. P1 is a protein which spans the inner membrane and studies have suggested P1 and P4 directly interact with each other [8]. The exact function P1 has on phage assembly is unclear, the largest portion of P1 exists in the cytoplasm and the interactions with host thioredoxin suggest that P1 is involved in the process of removal of P5 from assembling P5DNA particles. The function of P11, its location, and stoichiometry in the assembly complex are also unclear. Experiments have shown that P11 is absolutely required for assembly. Thioredoxin is believed to help P1 remove P5 from the P5DNA complex.

Phage assembly can be divided into six steps as shown in Figure 7. The first step involves the formation of the assembly sites. An assembly site consists of 14 molecules of P4 which come together to form a cylindrical pore on the outer membrane. Less is known about how many P1 and P11 form one assembly site, and in our model we are assuming 14 molecules of each form an assembly sites. The second step involve five copies each of coat proteins P7 and P9 binding to the assembly complex. The third step in the assembly process is the interaction of the packaging signal, located on the P5DNA complex, with the assembly site to initiate phage elongation. The packaging signal, indicated in red in Figure 7, is the region of ssDNA where two complementary strands located on the same piece of ssDNA bind together to form a double-strand hairpin. The packaging signal is not believed to be covered by the single-stranded binding protein P5.

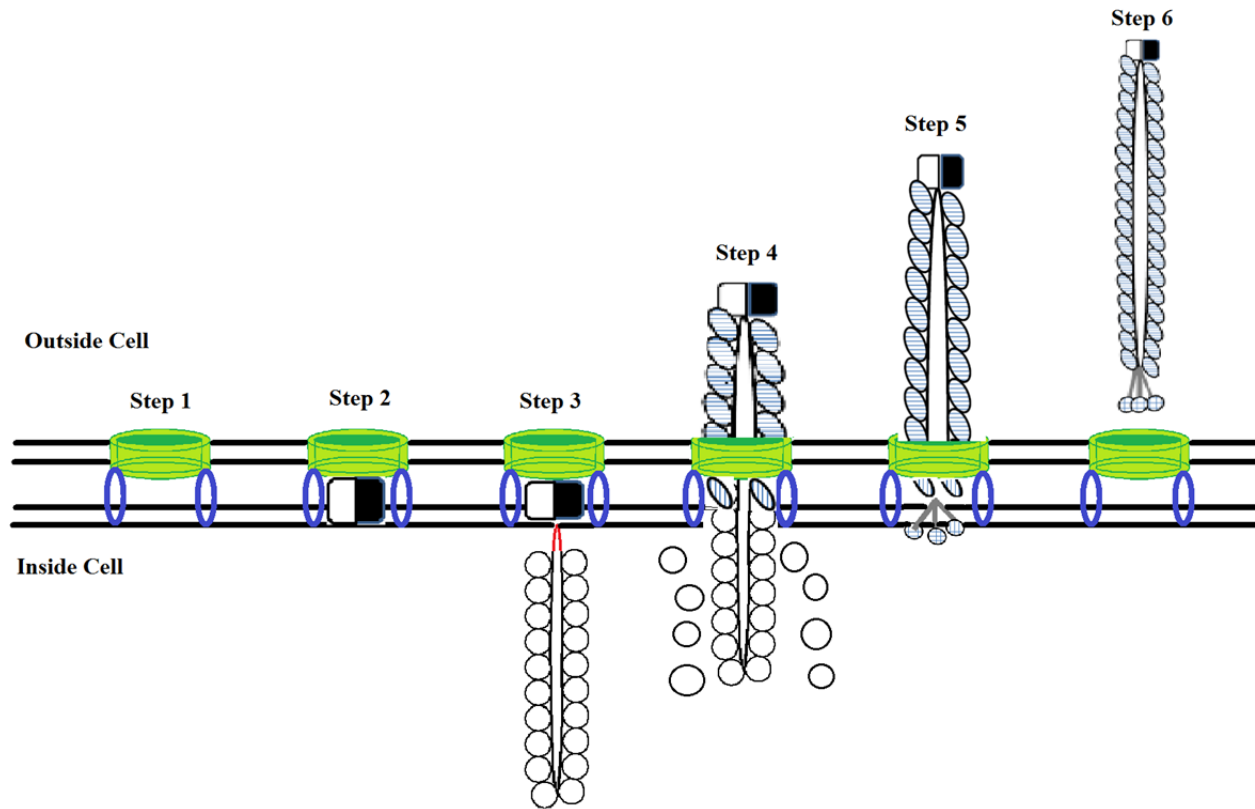






Figure 7 – A visual representation of the ordered in phage assembly. The figure depicts the six steps in phage assembly described in the text.  - indicates the P4 homomultimer complex located in the outer membrane,  - indicates the P1 and P11 associated with the assembly site,  - indicates the P7 and P9 complex associated at one end of the phage particle,  - indicates the P3 and P6 complex .

The fourth step involves the replacement of the P5 molecules with P8 molecule aided by the from ATP hydrolysis. During this step, the ssDNA is being exported outside of the cell as the major coat protein P8 is being attached to the nascent phage. The elongation/extrusion process proceeds until all of the P5 has been removed and replaced by P8 and the majority of the ssDNA has been passed out of the cell. The P5 molecules which are removed are recycled back into the cytoplasm. The fifth step is the termination of phage assembly which occurs when the end of the

ssDNA is reached. Termination involves the attachment of phage proteins P3 and P6. In the absence of P3, phage assembly cannot terminate and another P5DNA can bind to the assembly site to re-initiate step 3 [7]. The result will be phage particles containing multiple ssDNA genomes packaged end to end in long filaments that remain connected to the cell membrane. The final step is the regeneration of the assembly site and the release of the nascent phage to the extracellular environment. It should be noted that all of the coat proteins (P7, P9, P8, P3, and P6) are associated with the membrane before they are packaged into the phage particle. In order to clarify the depiction of phage assembly in Figure 7 the unincorporated coat proteins in the membrane were left out in the diagram.

1.4 Biotechnological Applications of Bacteriophages

Shortly after the discovery of bacteriophages in 1917 [33], it was postulated that they could be used as a treatment for bacterial infections in humans, called phage therapy. Phage therapy is accomplished by using a lytic bacteriophage (one which kills the host cell) to selectively infect harmful bacteria. However, research in phage therapy declined in the western world by the 1940s because of the discovery and success of the first antibiotic, penicillin [34]. Phage therapy was actively developed and used in the former Soviet Union and interest has revived in recent years with increased prevalence of antibiotic resistance in pathogenic organisms.

Antibiotics have been used successfully to treat both human infection and to prevent bacterial contamination in food but are becoming less effective due to antibiotic resistant bacteria [35]. Bacteriophages are a potential alternative and have been shown to be successful in reducing known bacteria contributing to food poisoning such as *E. coli O157:H7*, *salmonella*, and *Listeria monocytogenes* [35]. In 2006, the United States Food and Drug Administration (FDA)

approved the first bacteriophage treatment in the food industry, called List-Shield, to reduce *L. monocytogenes* in ready to eat food [36].

An alternative application, called phage display, was introduced in 1985 and it takes advantage of the bacteriophages' innate ability to replicate in large quantities [37]. Phage display involves selectively inserting randomized gene sequences into the viral genome at location coding for coat proteins [38]. The aim is to produce a protein which retains the ability to be assembled into a bacteriophage particle but displays a peptide, protein, or antibody fragment from the surface of the bacteriophage with the ability to bind to a target substrate, see Figure 8.

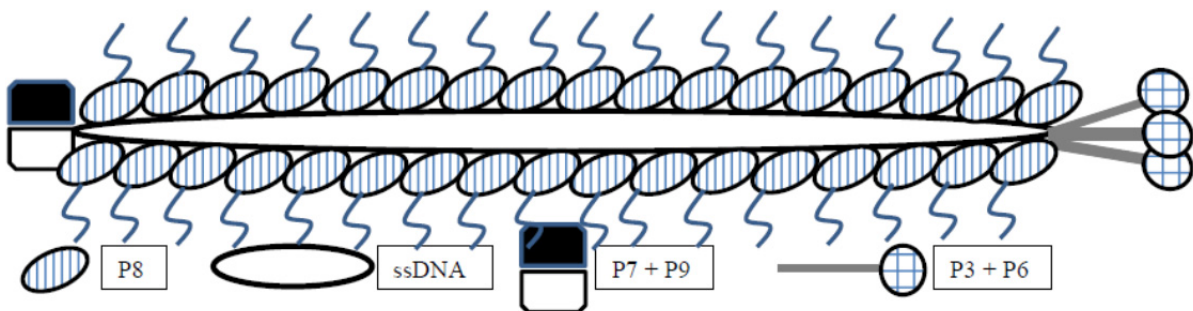


Figure 8 - A schematic diagram of the M13 bacteriophage structure with a gene modification in P8 which displays a peptide, protein, or antibody fragment on the external surface of the phage.

The M13 bacteriophage is currently being exploited for many novel applications using phage display. Phage display derived peptide and proteins are being employed in the assembly of materials [39], as a sensor [40], for drug discovery and delivery [41], and can be used to target and aid in imaging of cancer cells [42]. An example of a material made of phage particles is the development of nanowires. The phage genome coding for P8 can be modified to display a tetraglutamate peptide[43]. Metal ions can then bind to the tetraglutamate and be reduced to cover the phage species with a metal shell to form small nanowires. The main advantages to

using phage as a template for nanowires are that the processing steps do not require extreme operating conditions (high temperature and pressure) to create wires and that the phage particles allow somewhat controlled creation of wires at the nanometer scale [39].

1.5 Why Model Biology

The M13 bacteriophage kinetic model described in this thesis was built using information accumulated through several decades of research on the biology of M13. The literature on M13 is large and the biology of the life cycle is complex, therefore it is possible for inaccurate conclusions to have been drawn from false-positives or false-negatives in experiments [44]. The inaccurate conclusions may not be obvious until studying phage infection in a more holistic approach, such as a kinetic model. Kinetic models can be used to synthesize the fragmented information and test the consistency of experimental data and reveal gaps in understanding of the biological processes. Beyond testing the constancies of current understanding, the model can be employed to test complex hypotheses and guide future experimental designs [45].

Several types of models exist which require various degrees of detail/information to create and which allow different types of conclusions to be drawn from them. Diagram models are commonly used as visual aids and as tools to formulate new hypotheses. Although diagram models are useful for the understanding of a process, they lack the ability to make predictions. Boolean Models can be applied to study how networked genes interact if the genes have only two states, on or off [46]. Both diagram and Boolean models have been useful to study the qualitative behaviors of a system but neither provide quantitative descriptions [45]. Stoichiometric network analysis and kinetic modeling can both provide a quantitative description of a biological process. In a stoichiometric network analysis each biological process is set up as a chemical reaction and the result is evaluated under steady-state conditions [47]. Kinetic models

are sets of chemical reactions modeled by differential equations. They are usually the most difficult to create because each differential equation requires a kinetic parameter and many parameters have not been determined experimentally. Kinetic models are also the most descriptive and have the potential to give the greatest insight into a biological process [45].

1.6 Examples of Previous Virus Life Cycles

The T7 bacteriophage models developed in the laboratory of Jon Yin are an excellent example of the process of development and applications of a deterministic kinetic model to the life cycle of a virus. These T7 models were a major influence on the present work. The T7 viral infection process and life cycle are very different from the process described for M13. T7 completely takes over the host cell resources and replicates through a lytic cycle. The initial kinetic model development consisted of gathering the relevant information pertaining to the T7 infection and translating it to a set of ordinary differential equations [48]. The T7 model accounted for viral DNA entry, mRNA synthesis, protein synthesis, protein decay, DNA replication, phage assembly, and the inactivation of *E.coli* and T7 RNA polymerase. Once the model was created and validated, a set of in silico experiments were performed using the model. These experiments included testing for potential antiviral strategies [49], determining how random genome perturbations in both unlimited [50] and limited [51] host resource environments affect phage replication, inferring/predicting potential regulatory function of the infection process [52], testing how host cellular resources affect the rate of viral replication [53], and redesigning the genome to provide a simpler model that maintains key features of the infection cycle [54].

Another example of a viral model is the vesicular stomatitis virus (VSV) which accounts for the transcription, translation, and replication processes [55]. The VSV viral model did not

account for the entire replication cycle, however, it did use stochastic chemical kinetics to study how the intrinsic fluctuations of viral protein production early in the infection affect the overall dynamics of viral growth. The model was used to test all possible genome permutations [56], since VSV only contains 5 viral genes.

1.7 Motivation Behind Building a Kinetic Model of M13

This/our work involved the integration of the work of many groups carried out over the past 60 years of M13 bacteriophage research in the development of a kinetic model. The model sets up a set of ordinary differential equations that simulate DNA replication, transcription, mRNA processing and degradation, translation, phage assembly, protein-protein interactions, and protein-nucleic acid interactions. The model applies directly to the commonly utilized phage f1, fd, and M13 and can be expanded to model the behavior of more than 20 more related filamentous phage infections, for example filamentous phage Ike. Since M13 coexists with its host cell, it is assumed to have limited amount of cellular resources such as DNA polymerase 3 (DP3), RNA polymerase (RNAP), and ribosomes (R). The nucleoside triphosphates and amino acids were assumed to be unlimited. Experimental kinetic parameters were used when available, however many parameters were not available and were fit to experimental results. The simulation begins by assuming the cell just divided and a viral-single-stranded DNA (ssDNA) has entered the cytoplasm. The simulation ends when the infected cell is ready to divide at 60 minutes post infection.

The motivation behind building a kinetic model is to gain a better understanding of the biology of the phage, which is a technologically important system. Improved understanding is expected to translate into improved platforms for phage display and phage material synthesis by serving as a design tool to help refactor the M13 genome into a more easily manipulated

sequence. The natural phage genome sequence is tightly packed together with overlapping protein sequences and complex control elements. Greater control over the infection process can accelerate the research involving novel applications of phage display. These novel applications include: the development of phage-based systems for sensing biochemical agents in the detection of human and food-borne pathogens; utilizing the self-assembly properties inherent in the phage structure to utilize phage as a template for nanotechnology, such as nanowires; and a vehicle to distribute gene therapy drugs to specific targets [57].

Although ambitious, we chose to develop a kinetic model for the M13 virus life cycle due to the extensive knowledge surrounding the viral infection and the potential engineering applications that have been demonstrated, such as the use of phage display for material synthesis. The dynamics of viral growth/infection was studied through the formation of a kinetic model utilizing ordinary differential equations. The model was able to predict many phenotypic behaviors observed in wild-type phage infection. It was also used to point out gaps in the current knowledge.

Chapter 2: Model Description

2.1 Summary

The model of the phage life cycle simulates the replication of phage DNA, transcription and degradation of phage RNA, translation of phage proteins, and production/assembly of progeny phage particles. The model simulation begins when the ssDNA is separated from the phage capsid and is introduced into the cell, the process of phage infection is not explicitly included in the model. All of the coat proteins of the infecting phage are assumed to be inserted into the cellular membrane and one ssDNA genome is inserted into the cytoplasm. At the start of the simulation, 20 percent of the total host enzymes DNA polymerase 3 (DP3), RNA polymerase (RNAP), and ribosomes are assumed to be available to the phage for replication [58]. The allocation of cellular resources to the phage is based on experimental observations of cellular growth retardation and estimates of the average physiological burden of infection [59]. The amino acid and nucleic acids supplied by the host cell are assumed to be abundant and do not limit the rate of phage replication.

The initial values for almost all model species are zero at the start of the infection except for the coat proteins, ssDNA, and host enzymes. The five coat proteins which are inserted in the membrane during the infection process are assumed to retain the ability to function in particle assembly. The initial amounts of phage proteins in the inner membrane are 2700 P8 molecules and 5 molecules of each P3, P6, P7, and P9. The ssDNA has an initial value of 1 molecule because we assume a single phage genome is delivered by the infecting particle. Filamentous phage infections produce somewhere between 1% and 5% of particles with two packaged genomes [7]. Finally, the host enzymes have initial values which are 20 percent of the average number of host enzymes observed in actively growing *E. coli* [58].

Table 3 – Displays the species name with the corresponding initial number of molecules present at the start of the infection. All other species not present in the table have a value of 0 molecules.

Species Name	Initial Value (Molecules)
P8	2700
P3	5
P6	5
P7	5
P9	5
ssDNA	1
RNA Polymerase (RNAP)	1280
DNA Polymerase III (RP3)	3
Ribosomes	7880

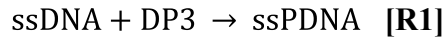
2.2 Formation of Viral dsDNA

The current section discusses the biology of dsDNA synthesis, the simplifications made to model the process, and how it was modeled. The synthesis of dsDNA requires a complete single-stranded DNA (ssDNA) to be synthesized or inserted into the cell. Since the conversion of the viral ssDNA to dsDNA is the first step in the infection process following infection, we will discuss the formation of dsDNA synthesis first followed by ssDNA synthesis.

The conversion of ssDNA to dsDNA is a complex process which involves many host enzymes and resources. Upon entering the cell, the viral ssDNA is immediately converted to supercoiled double-strand DNA called replicative form 1 (RF1) by a set of *E. coli* host proteins [9]. Conversion of ssDNA to RF1 occurs by the ordered actions of *E. coli* enzymes RNA polymerase, DNA polymerase III (DP3), DNA polymerase I, and DNA ligase. RNA polymerase is the first enzyme to bind to the ssDNA and creates a short RNA primer. DNA polymerase III (DP3) will bind to the primed ssDNA and synthesize the complementary strand. Next, DNA polymerase I will remove RNA primer and replace it with the complementary DNA strand. Finally, DNA ligase joins the strands together to complete the circle. The model simplifies the complex process of dsDNA formation into two steps, that involves the binding and elongation of

DNA polymerase 3 (DP3) to the single-stranded template. The justifications to the model simplifications are discussed in the following paragraphs.

The rate of DNA polymerase 3 (DP3) binding is likely limited by the available number of molecules the host cell synthesizes. Measurements of DNA polymerase 3 (DP3) levels in actively growing *E. coli* indicate approximately 10-20 copies per cell [60]. Complimentary strand synthesis of phage DNA is further limited by the allocation of only 10-20% of the total host cell machinery [58], which reduces the available number of DP3 for phage infection to 1-4 copies per cell. The model assumes 3 RNAP are available for phage infection at the beginning of the infection process because the average was 2.5 molecules of DP3 (of the 1-4 copies of DP3 allocated) and the number was rounded up. The binding of DP3 available for replication is accounted for by the following binding reaction R1, where ssPDNA is the ssDNA/DP3 binding complex. The rate expression for R1 follows simple second order mass action kinetics, E1:



$$\text{Rate} = C_1 * \text{ssDNA} * \text{DP3} \quad [\mathbf{E1}]$$

An experimental determination of the rate constant C_1 could not be found through literature. Therefore we chose a rate of C_1 be $0.1 \text{ (molecules second)}^{-1}$ because the rate of binding did not limit the rate of dsDNA formation. Rather the availability of the DP3 enzyme and the rate of elongation determined the rate. Due to the insufficient information on DP3 binding, the previous assumption could be wrong when more experimental evidence is determined or found through literature. The inclusion of the binding step of the DNA polymerase 3 is essential in the model because the number of DP3 enzymes is a limiting factor in DNA replication. Once DP3 is bound to the ssDNA, which we represent by the formation of the model

species ssPDNA, elongation initiation is assumed to occur instantaneously and the rate of formation of RF1 is limited by the rate of DP3 elongation. The rate of elongation follows first order mass action kinetics and is described by reaction and rate equations R2 and E2:



$$\text{Rate} = C_2 * \text{ssPDNA} \quad [\text{E2}]$$

The final products for R2 are the recycled DP3 enzyme, RF1, and 5 promoter sites (DA, DB, DH, DZ, and DW). The 5 promoter sites are assumed to be independent species within the model, despite being physically located on RF1 DNA. We chose to decouple the promoter sites from the DNA to make transcription easier to model because each promoter sites is considered an independent species, and the rate of transcription is assumed to be unchanged in the nicked (RF2) and un-nicked (RF1) DNA. The decoupling of the promoter sites from RF1 and RF2 imposes the following conditions on the model to remain biologically consistent:

$$[\text{RF1} + \text{RF2}]_{\text{total}} = [\text{DA}]_{\text{total}} = [\text{DB}]_{\text{total}} = [\text{DH}]_{\text{total}} = [\text{DZ}]_{\text{total}} = [\text{DW}]_{\text{total}}$$

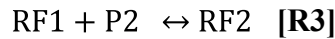
The first order rate constant C_2 was calculated using the average rate of DNA Polymerase 3 (DP3) elongation and the length of the genome. The rate of DP3 elongation was reported to be 750 nucleotides/second in *E. coli* [60]. We calculate C_2 assuming DP3 does not pause or reverse on the genome. The length of the phage genome in our model is taken to be 6407 nucleotides long which is consistent with the three closely related filamentous phage species f1 (6407 Nucleotides), m13 (6407 Nucleotides) and fd (6408 Nucleotides). A first order rate constant was obtained by dividing the rate of elongation by the length of the nucleotide sequence of the genome, see calculation of C_2 below:

$$C_2 = \frac{\text{Rate of DNA Polymerase Elongation}}{\text{Length of Phage Genome}} = \frac{750 \frac{\text{Nucleotides}}{\text{Second}}}{6407 \text{ Nucleotide}} = 0.117 \frac{1}{\text{Second}}$$

According to the rate constant C_2 , once DNA polymerase 3 (DP3) has bound to the ssDNA, the completion of the RF1 DNA synthesis takes 8.55 seconds. The model is assuming host enzyme DNA gyrase can supercoil the DNA at a rate much faster than the rate of DP3 elongation (which may or may not be correct). The model currently assumes the rate of dsDNA synthesis is limited in the available number of DP3 enzymes and the rate of elongation. However it is still possible that the rate of DNA polymerase III binding is the rate limiting step.

2.3 P2 Nicking RF1 DNA

Once the supercoiled RF1 DNA has been synthesized, transcription quickly occurs followed by translation of all of the phage encoded proteins. The details of phage gene transcription and protein translation are discussed after the phage DNA replication section despite transcription and translation occurring first before phage dsDNA replication begins. The replication of phage DNA will not occur until viral protein P2 is translated. Viral protein P2 is required to nick the RF1 DNA to produce replicative form 2 DNA (RF2). RF2 is a relaxed dsDNA with a break in the (+) strand, with no superhelical turns, and a single molecule of P2 bound [9]. The nicking of RF1 to produce RF2 is modeled by the following reaction:



R3 is a simplification of the process depicted in Figure 5, which is thought to be a more accurate representation of the biology. Figure 5 shows two molecules of P2 binding to RF1 DNA to convert it to RF2₁. RF2₁ indicates a relaxed dsDNA with a break in the (+) strand, with no superhelical turns, and with two molecules of P2 bound. Two outcomes were shown to occur once the two P2 molecules have bound to it. First it showed both molecules of P2 dissociated

forming RF4, which is relaxed double-stranded DNA with no nicks in the positive strand. Second, only one P2 dissociates to form RF2, the active form of dsDNA to which DNA polymerase 3 (DP3) can bind to initiate ssDNA synthesis. The complete details behind P2 binding and unbinding are not clearly understood and the rates of P2 binding and unbinding are unknown. The model has been simplified to account for only one molecule of P2 binding to RF1 DNA. The simplification, as seen in Figure 9, assumes only one P2 molecule is needed to nick RF1 DNA. Additionally, we assume RF4 DNA (relaxed, dsDNA, with no superhelical turns, and no nicks in the positive strand) will not accumulate in the cell because DNA gyrase is assumed to act sufficiently fast to convert RF4 DNA to RF2 DNA. Therefore, we do not account for the RF4 species in the model.

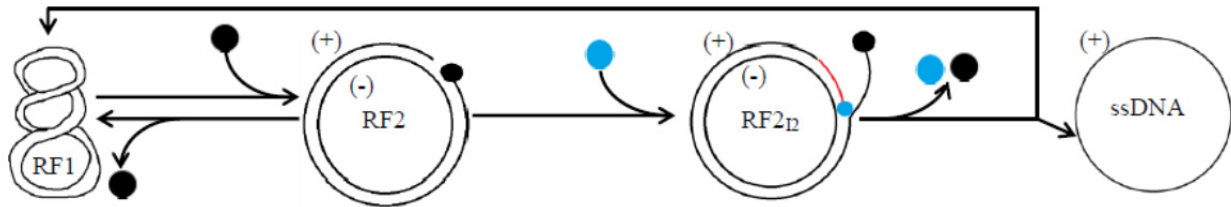


Figure 9 – A visual depiction on how the biology is modeled, which is an simplified rolling circle DNA replication mechanism. ● - Represents one P2 molecule, ● - represents the DP3 and replicative helicase complex, and the red indicates the newly synthesized (+) strand.

The rate expression, E3, models the association and dissociation of P2 on dsDNA and follows simple mass action kinetics. The equation consists of two unknown variables C_3 and C_4 . C_3 is a second order rate constant which describes the rate at which P2 nicks DNA. C_4 is a first order rate constant and describes the rate of dissociation of the P2/DNA (RF2) complex.

$$Rate = C_3 * RF1 * P2 - C_4 * RF2 \quad [E3]$$

Constants C_3 and C_4 have not been measured experimentally but they can be related to each other to reduce the number of unknowns in the model by one. Constants C_3 and C_4 are related through the observed steady-state behavior of phage dsDNA pool measured in an in-vitro study. Measurements from the in-vitro study concluded that 60% of RF1 DNA is nicked to form RF2 and 40% forms RF4 [61]. RF4 is a relaxed dsDNA with no break in the (+) strand and no superhelical turns [9]. We are assuming the percentage of dsDNA in the RF2 form in-vitro is equivalent to in-vivo results but there is a high uncertainty in the assumption. Also we are assuming the percentage of RF4 observed in vitro to be equivalent to the sum of the in-vivo percentage of RF1 and RF4 DNA. It is the sum because in an in-vivo setting DNA gyrase will be present to convert it to RF1 DNA. Since our model does not take into account RF4, 40% of the total DNA can be assumed to be in the RF1 DNA and 60 percent assumed to be RF2 DNA. Additionally, the production of P2 is shown to reach a plateau at about 1500 molecules per cell and the accumulation stops at approximately 20 minutes post infection [62]. Assuming E3 is at steady-state, applying the information of the in-vitro study on dsDNA, and knowing the steady state of P2, the C_3 and C_4 dependence is calculated by the following steps:

$$0 = C_3 * RF1 * P2 - C_4 * RF2$$

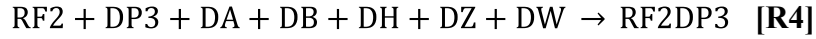
$$C_3 * RF1 * P2 = C_4 * RF2$$

$$\frac{C_3}{C_4} = \frac{RF2}{RF1 * P2} = \frac{0.6 * DNA_{total}}{0.4 * DNA_{total} * 1500} = 0.001 \frac{1}{s}$$

2.4 Production of ssDNA

Once the DNA has been converted to RF2, DNA polymerase 3 (DP3) and helicase can bind to RF2 to initiate DNA replication. Helicase will separate the (+) and (-) strands while DP3 uses the (-) strand as a template to create a new (+) strand while displacing the old strand. The

old strand is the nascent ssDNA produced in rolling cycle DNA synthesis. The nascent ssDNA can be converted to either RF1 DNA or P5DNA depending on the progress of the infection. DNA ligase will close the nick on the dsDNA and P2 will re-circularize the displaced ssDNA. The model accounts for the binding and elongation of DP3 in the same manner as in the conversion of ssDNA to RF1. Reaction equations R4 and R5 describe the binding and elongation by DP3 respectively. P2, RF1 and DP3 are regenerated in equation R5 after the elongation has been complete.



When DP3 attaches to RF2, transcription is assumed to be disrupted momentarily while DP3 is elongating, thus consuming the promoter sites DA, DB, DH, DZ, and DW. Once DP3 has produced a new ssDNA, the promoter sites are regenerated to allow an RNA polymerase (RNAP) to attach to the promoter sites. The model does not allow RNAP to initiate transcription once DP3 has bound to dsDNA. Additionally, the model does not address what should happen to the current actively transcribing RNAP on the dsDNA undergoing DNA replication. Therefore, we are currently assuming the possible negative effects of DP3 and RNAP interactions are minimal and are not accounted for in the model.

Reaction R4 describes binding of RF2 and DP3, and follows second order mass action kinetics. The reaction rate equations, R4 and R5, are modeled as simple mass action kinetics as described by equations E4 and E5. We assume DP3 has the same binding affinity for ssDNA and dsDNA (E1 and E4 have the same rate constant) because no information has been found to support the statement, and the rate of DP3 elongation is the same on each type of DNA (E2 and

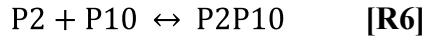
E5 have the same rate constant). Variations in the rate of binding and elongation of DP3 probably exist but we are assuming the extent of the variation is minimal but a large variation is possible.

$$Rate = C_1 * RF2 * DP3 \quad [E4]$$

$$Rate = C_2 * RF2DP3 \quad [E5]$$

2.5 P10 Inhibits DNA Synthesis and Phage Replication

Phage replication is observed to be inhibited when the number of copies of P10 exceeds the number of P2 copies in the cell [14, 15, 63]. The inhibition is thought to occur by P10 directly inhibiting P2 function, which prevents RF2 formation and DNA synthesis. We proposed a binding reaction between P2 and P10 to form a P2/P10 complex (P2P10). The binding stoichiometry of P2/P10 is assumed to be 1 to 1, as shown in the reaction equation R6.



To the best of our knowledge, no experimental evidence has directly suggested P2 and P10 bind together. Although speculative, we believe the interaction of P2 and P10 to be a reasonable mechanism through which P10 could affect P2 function. The coding region for P10 is completely contained within the reading frame of P2. In other words, P10 has the exact sequence as the last 112 amino acids in P2. Additionally, P2 is thought to be in the form of a dimer but it has not been quantified [9]. If P2 can form a dimer with itself and P10 contains a similar amino acid sequence, then P10 may retain the region of the protein required for binding.

The equation describing P2-P10 binding, R6, follows simple mass action kinetics where both binding (C_5) and dissociation (C_6) parameters are unknown. Since no experimental evidence exists about the binding and dissociation the binding constant were chosen somewhat arbitrarily.

The association constant was chosen to be $1 \text{ (Molecule * Second)}^{-1}$ and the dissociation constant was chosen to be $10^{-15} \text{ (Second)}^{-1}$. With the constants chosen, if $P2 > P10$ then all P10 will be bound a single P2 leaving no free P10 in the model. If $P2 < P10$ then no free P2 will be in the system. If P2/P10 binding is the mechanism through which P10 affects P2 function, then the high binding affinity probably is not correct since the P2P10 binding complex has not been observed. Additionally a proposed mechanism discussed in the paper would require free P10 therefore the current constants are place holders until experimental evidence is found. However, since free P10 currently has no function in the model other than binding to P2, the improper disposition of P10 does not affect the outcome of the model. With such a strong binding affinity, when the number of molecules of P2 equals P10 DNA replication will stop immediately.

$$\text{Rate} = C_5 * P2 * P10 - C_6 * P2P10 \quad [\text{E6}]$$

The reverse reaction is left in the model because it is an assumption that will be relaxed if P2P10 is experimentally confirmed and more information is found. P10 has been shown to be required for P5DNA formation to occur [14, 15] and is likely to be needed separately from the P2P10 form. A proposed mechanism which will incorporate more of the known biology into the model is discussed in chapter 3 and requires free P10 to accumulate in the model system.

2.6 Hill Equation Overview

The following section and two other sections in the description of the model (“Translation” and “Assembly of New Phage”) multiply a hill equation into the rate equations. A hill equation is a way of describing a system which can have two states, on or off. Hill functions come in multiple forms that contain three parameters; a factor X, usually a concentration of a species in a reaction, a constant K_m , which is an equilibrium constant describing the midpoint of the transition, and the exponent n, which determines the steepness of the transition. The on and

off states can be described numerically as 1 or 0. Therefore, the hill coefficient can take on any value between 0 and 1 depending on the value of X in comparison to K_m .

Two versions of the hill coefficient exists describing the increasing or decreasing nature of the hill equation as a function of X. The positive interaction will cause the function to increase from 0 to 1 as X increases, called an increasing hill function (IHF). The negative interaction will cause the function to decrease from 1 to 0 as X increases, called a decreasing hill function (DHF). The exponent usually takes on an integer value between 1-4 and affects the rate at which the function output changes.

$$IHF = \frac{X^n}{X^n + K_m^n} \quad \text{and} \quad DHF = \frac{1}{1 + (X/K_m)^n}$$

To understand what a hill function is doing in the rate equation it is easiest to look at an example. Let's assign a value for $K_m = 250$ and X to a range of values between 1-2000 for the increasing hill function. The hill function will be plotted 4 times varying n between 1-4, see Figure 10 (left). Looking at the Figure 10 (left), when X is equal to zero, the function also equals to zero. When X is equal to 250 (which is equal to K_m), the value of the equation is 0.5 and is independent of the value of n. When X is greater than K_m the rate at which the function increases to 1 is dependent on n. When $n = 1$, the function takes on a value of 0.9 when $X = 2000$ and when $n = 4$ the function takes on a value of 0.9 when X is ~ 400 . Therefore n dictates the steepness of the slope. The characteristics of the increasing hill function also apply to the decreasing hill function except the values go from 1 to 0.

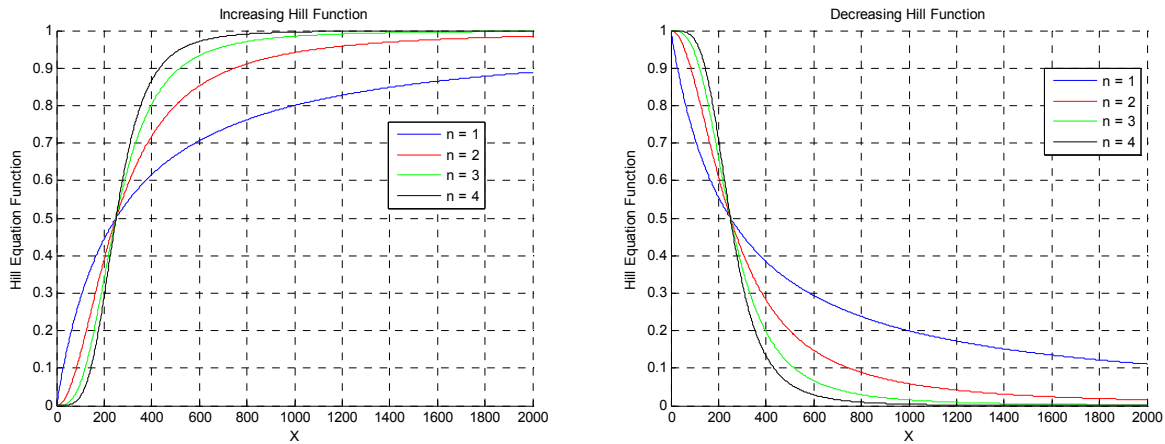


Figure 10 – (Left) the output of an increasing hill function when $K_m = 250$ and X is varied between 1-2000. Each line represents the output of the hill function when n is equal to 1, 2, 3, and 4. (Right) the output of a decreasing hill function when $K_m = 250$ and X is varied between 1-2000. Each line represents the output of the hill function when n is equal to 1, 2, 3, and 4.

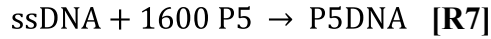
The value of n describes the extent to which the process is cooperative. An value of n equal to 1 describes a non-cooperative process and any n value greater than 1 describes a cooperative process. The discussion about the hill function in this paper was aimed to familiarize the reader with the functionality of the hill equations. A more detailed overview of the hill functions and history of the use of the hill function in kinetic modeling can be found in [64].

2.7 Shift from RF1 to P5DNA Production

Early in the infection process the single-stranded DNA produced by rolling circle replication is converted to double-stranded forms, building a pool of replicative DNA. As the production and level of phage protein P5 increases during the course of infection, ssDNA is sequestered and ultimately packaged in new progeny phage. The replication cycle of phage dsDNA replication occurs in three steps. First RF1 is nicked by P2 to produce a relaxed dsDNA with a nick in the (+) strand. Secondly, DNA polymerase III binds to RF2 DNA to synthesize the (+) strand using the (-) strand as a template. The old (+) strand is displaced and the result of the second step is RF1 DNA and ssDNA. Lastly the ssDNA is converted to RF1 DNA by host enzymes. Early in the infection process, the dominant fate for nascent ssDNA will be the

conversion to RF1 DNA. As the virus infection progresses, the phage protein P5 will accumulate and ssDNA will be sequestered by approximately 1600 molecules of P5 [59, 65, 66]. We named the ssDNA sequestered by P5 molecules P5DNA. P5DNA cannot be converted to RF1 DNA and it is the form of ssDNA which can interact with the assembly proteins to initiate phage assembly.

The binding of P5 to ssDNA has been determined to be a cooperative process [67], meaning the binding of the first P5 molecules increases the binding affinity of the next P5 molecule. The precise series of events leading up to the formation of P5DNA will involve thousands of individual reactions if each P5 binding event is modeled. We have chosen to simplify the P5 binding process in our model and rather than include each individual P5 binding reaction we approximate the P5DNA formation in a single reaction, as described by R7.



Reaction R7 depends on the number of ssDNA and P5 molecules within the cytoplasm. An increasing hill function is applied to the rate equation E7 to ensure P5DNA synthesis does not occur until the experimental determined threshold of P5 molecules is reached. The switch has been shown to occur when approximately 1100 molecules of P5 are made [68]. The threshold does not indicate when P5DNA synthesis will become dominate over RF1 DNA synthesis but the point in the model where P5DNA synthesis can begin.

$$\text{Rate} = C_7 * \frac{P5^{n_1}}{k_{m1}^{n_1} + P5^{n_1}} \text{ssDNA} * P5 \quad [\text{E7}]$$

Experiments have shown at approximately four minutes post infection P5DNA synthesis will start and by 20 minutes post infection 95% of all nascent ssDNA is converted to P5DNA [13]. The previous experiment suggests the dominate fate of all nascent ssDNA does not change rapidly but gradually over 15 minutes. In other words, the synthesis of RF1 DNA and P5DNA

will occur simultaneously during the first 20 minutes post infection. In reality, RF1 synthesis can occur during the entire infection process but it is most likely at a much depreciated rate when compared to the first 20 minutes. The literature suggesting a gradual change from dsDNA to P5DNA synthesis led to our choice of the hill coefficient, n_1 , to be equal to one in the rate equation for P5DNA formation, indicating a non-cooperative process.

Choosing a hill coefficient of one indicates the process is non-cooperative. A cooperative process would have a hill coefficient greater than or equal to two. We are not assuming P5 binding to ssDNA is not cooperative but we are assuming the overall process of one ssDNA being converted to P5DNA can be described as a non-cooperative process. In other words, describing the individual reactions of P5 binding to ssDNA is cooperative but when describing P5DNA formation in a single step the overall process it is not cooperative.

Reaction R7 is a significant simplification of the binding of P5 to ssDNA and it most likely does not represent all of the events leading up to P5DNA formation. For example, our model does not consider the binding of P5 to ssDNA with an RNA primer. There is no particular reason to think that partially bound ssDNAs do not exist early in the infection process. Very early in the infection when the P5 concentration is relatively low, P5 should begin to bind to ssDNA cooperatively. However, the complete sequestering of ssDNA will be disrupted by DNA polymerase 3 (DP3) because the complete sequestering did not occur fast enough or not enough P5 was present for complete sequestration. The previous statements assumes DNA polymerase III (DP3) will be unaffected by bound P5 molecules. The assertion that DP3 can remove P5 is supported by experimental evidence that in the presences of high levels of P2 ssDNA is converted to RF1 regardless of P5 levels [14, 15].

Rate equations E1 and E7 both compete for ssDNA. Rate equation E1 describes the binding of DP3 to ssDNA which initiates RF1 DNA synthesis. Rate equation E7 describes the sequestration of ssDNA by P5. Early in the infection when P5 levels are low, the rate of E1 will be greater than E7. However as the reaction progresses and P5 accumulates in the cell, E7 will become the dominant pathway. If the rate of P5 and ssDNA accumulation remains constant in the model, the time at which E7 becomes dominant over E1 can be controlled by varying the rate constant C_7 . Decreasing C_7 will result in an increase in dsDNA synthesis because P5DNA synthesis will be slower. Increasing C_7 will result in a decrease in dsDNA synthesis because P5DNA conversion occurs faster. The rate constant C_7 was systematically varied until the model matched the experimentally determined number of dsDNA (approximately 50 strands of DNA). In equation E7, K_{m1} is the number of P5 molecule needed to begin the sequestering ssDNA, and n_1 is the hill coefficient chosen to be 1.

Table 4 – The description, symbol, value, and source for the parameters used in the DNA replication. Parameters with reference of N/A indicate a parameter that was fitted to observed experimental results. #indicates the relative ratios of the two parameters were estimated but the magnitude of the value is unknown.

Parameter Description	Symbol	Value	Reference
Replication			
DNA Polymerase III Binding Rate to ssDNA and dsDNA	C_1	$0.1 \text{ (Molecule*Second)}^{-1}$	N/A
DNA Polymerase III Elongation Constant	C_2	$0.117 \text{ (Second)}^{-1}$	[60]
P2 Nicking RF1 DNA to Form RF2 DNA	C_3	$1 \text{ (Molecule*Second)}^{-1}$	[61]#
P2 Dissociation to RF2 DNA	C_4	$1.0E2 \text{ (Second)}^{-1}$	[55]#
P2/P10 Association Constant	C_5	$1 \text{ (Molecule*Second)}^{-1}$	N/A
P2/P10 Dissociation Constant	C_6	$10^{-15} \text{ (Second)}^{-1}$	N/A
Rate of P5 Sequestering ssDNA	C_7	$1E-3 \text{ (Molecule*Second)}^{-1}$	N/A
Threshold of P5 Molecules for P5DNA Formation	K_{m1}	1100 (Molecules)	[68]
Hill Coefficient for P5DNA Formation	n_1	1 (Unitless)	N/A

2.8 Transcription

The levels of phage proteins in the infection process are controlled in a highly complex manner. At the level of the genome, a set of nested transcripts produces more copies of mRNA for the most abundant proteins. In other words, multiple overlapping polycistronic transcripts are

produced such that all of the transcripts encode the most abundant proteins. The multiple overlapping primary transcripts produced are processed and degraded at different rates to produce a larger number of secondary mRNA species. At the mRNA level, protein production is controlled both by the strength of ribosome binding sites and by coupling of translation of genes in polycistronic transcripts. Further control over the translational efficiency of various genes is exerted by the level of the phage encoded single-strand DNA binding protein P5. P5 affects the translational efficiency by binding to specific species and inhibiting translation.

In vitro studies have been able to reveal a total of 8 promoters within the phage genome [21]. There are five promoters in the more frequently transcribed region (coding for viral proteins P2, P10, P5, P7, P9 and P8) and 3 in the less frequently transcribed region (coding for viral proteins P3, P6, P1, P11, and P4). In vivo studies have only been able to observe 3 active promoters in the more frequently transcribed region [24]. All of the promoters in the frequently transcribed region terminate at a single location distal to gene 8. The three promoters in the frequently transcribed region observed in vivo correspond to the three strongest promoters observed in vitro [21]. The infrequently transcribed region is transcribed by two different promoters and terminators. A very weak promoter, which had not been detected in vitro, is believed to lie within the rho-independent terminator used for the frequently transcribed region [27]. The transcript ends at the rho-dependent terminator distal to gene 6. This promoter produces a transcript that codes for phage minor coat proteins P3 and P6 as well as phage assembly proteins P1 and P11. Gene 4 is transcribed by its own promoter and terminated by one of four different rho-dependent terminators that lie within the intergenic region [20].

In vivo studies suggest that genes 1 and 11 are part of an operon that also encodes genes 3 and 6 [6] despite the rho-dependent terminator separating the two pairs of genes. The rho-

dependent terminator has been shown to be only 60-70 % efficient at terminating RNAP elongation [5]. Therefore, genes 3 and 6 are only transcribed from RNAP initiated by promoter DZ. Looking at the genome in Figure 3 suggests P4 would also be in the same operon because the next terminator is in the intergenic region, which is distal to P4 coding region. We assume P4 is not a part of the P1, P11, P3, and P6 operon because the longer operon containing P4 is not observed [6]. The failure to observe a transcript long enough to contain genes for P3, P6, P1, P11, and P4 may also indicate a very fast processing event. P4 contains its own promoter site and a transcript coding only for P4 has been observed. We expect most of the translation of P4 comes from the initiation of transcription at the promoter site DW.

Table 5 lists the names of each promoter site, the mRNA species that arises from initiation at the promoter site, which proteins are encoded by the mRNA, the nucleotide length of the mRNA transcript, and the relative promoter site strengths. The strongest promoter site was experimentally determined to be DB. The relative strength of a promoter site indicates the frequency at which an RNAP binds to the open promoter site in comparison to other promoter sites. The promoter sites in Table 5 are all compared to DB. For example, DA was determined to have relative promoter site strength of 0.6, meaning transcription initiation will occur 40 percent less frequently at DA than DB.

The values for the promoter strength in the more frequently transcribed region were taken directly from the in-vitro study and were all relative to the promoter strength of DB [21]. In the same study, the relative promoter strengths in the less frequently transcribed region were also measured relative to the most active promoter in the region, DW. In this study no promoter strength was evident for DZ because the transcript was not detected [21]. The promoter site strengths for promoters DZ and DW (values shown in Table 5) were not experimentally

determined but chosen to conform to two known conditions. DW and DZ are transcribed less frequently than the promoters of DA, DB, and DH. Secondly, DZ is expected to be weaker than DW because the activity of the DZ promoter site was not detected in vitro.

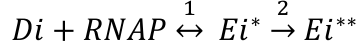
Table 5 – Each primary transcript created by M13 infection is displayed in the following table. The table lists the name of promoter site, the mRNA name, the proteins coded within the mRNA, nucleotide length and the relative promoter strength of each promoter [21]. * indicates values that were fitted to experimentally observed results.

	Promoter Site	mRNA Name	Proteins Coded	Nucleotide Length	Relative Promoter Strength
More Frequently Transcribed Region	DA	A	P2, P5,P7,P9 and P8	1800	0.6
	DB	B	P10,P5,P7,P9 and P8	1000	1.00
	DH	H	P9 and P8	320	0.65
Less Frequently Transcribed Region	DZ	Z	P3 and P6	1612	0.02*
	DZ	Y	P3, P6, P1, and P11	2663	0.02*
	DW	W	P4	1215	0.20*

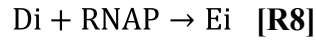
The model describes the transcription of the mRNA as a three step process within the cell. First, RNA polymerase (RNAP) binds to the promoter site (binding), secondly an RNA polymerase transcribes at a constant rate of elongation (elongation), and finally the RNAP and the mRNA are released from the DNA (termination). The first two steps are accounted for in the model while the rate of termination is assumed to occur instantaneously after the RNAP has reached the terminator [69, 70].

2.9 RNAP Binding to the Promoter Site

RNAP binding to the promoter site (D_i) can be further described as two separate steps, which consist of the binding of the RNAP making the pre-initiation complex (E_i^*), and the initiation of mRNA elongation (E_i^{**}) [69].



Our simplified model of transcription initiation assumes that the rate of RNAP binding to the promoter site is the determining factor of initiation. The rate of RNAP elongation initiation (~1 second) is faster than RNAP dissociation for all promoter sites on the phage genome (30 seconds – 50 hours) [69]. Therefore, we are assuming once RNAP binds to the promoter site, elongation will initiate instantaneously. R8 generically describes the process of initiation for all promoters *i* (DA, DB, DH, DZ, and DW) where *Di* is the promoter site and *Ei* signifies an RNAP bound to the promoter site.



The rate of RNAP binding to promoter site follows 2nd order mass action kinetics, where $C_{8,i}$ is dependent on the rate of RNAP binding to DB and the relative promoter strength RPS_i :

$$Rate = C_{8,i} * Di * RNAP \quad \textbf{[E8]}$$

$$\text{Where: } C_{8,i} = RPS_i * C_{8,B}$$

The rate of RNAP binding to the promoters sites on the M13 DNA genome have been measured and a range of binding constants calculated, $10^6 - 10^7 \text{ M}^{-1} \text{ s}^{-1}$ (where M is molar concentration or [Moles/Liter]) [69]. However, it was unclear in the paper on how each promoter site differed in binding. Therefore, the rate constants were converted from $\text{M}^{-1} \text{ s}^{-1}$ to $\text{Molecule}^{-1} \text{ s}^{-1}$, by assuming the mean cell volume is $7.5 * 10^{-16} \text{ L}$ [71] and using Avogadro's number ($6.022 * 10^{23} \text{ molecules/mole}$), see calculation below. The lower bound of the measured binding constant was used for the strongest promoter site B, $0.01 \text{ Molecule}^{-1} \text{ s}^{-1}$. The lower bound was chosen somewhat arbitrarily because the information on these parameters is scarce. The

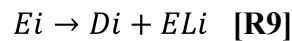
outcome of the model does not change significantly if the upper bound was chosen, indicating transcription is not limited by the rate of RNAP binding but rather by the number of available RNAP enzymes initialized at the beginning of the simulation.

$$10^6 - 10^7 \frac{L}{Mol * s} * \frac{1 Cell}{7.5 * 10^{-16} L} * \frac{1 Mol}{6.022 * 10^{23} Molecule} \approx 10^{-2} - 10^{-1} \frac{Cell}{Molecule * s}$$

The number of RNAP per cell in *E. coli* cells were measured to be between 1,500-11,400 molecules per cell [72]. Since phage infection only takes over ~20% of host cell metabolism, one can predict phage infection has a range of 300 – 2280 RNAP available for production of phage mRNA species. The model assumed a static number of 1,280 which is the average of the predicted amount of RNAP utilized by phage infection.

2.10 RNAP Clearing the Promoter Site

The model accounts for the possibility of multiple RNAPs simultaneously transcribing the same mRNA species on the same dsDNA. Therefore at some time after transcription initiation has begun the RNAP will be sufficiently far from the promoter site such that an additional RNAP can bind to the promoter. The RNAP clearing the promoter site will regenerate the promoter site (*Di* species) which allows it to bind to another free RNAP. The RNAP that recently cleared the promoter site has not finished transcribing the mRNA. Promoter site clearance is treated in the model through the generation of a new species, *ELi*, which denotes an elongating RNAP that does not occupy a promoter site. The reproduction of the promoter site follows first order mass action kinetics as described by the generic equation R9:

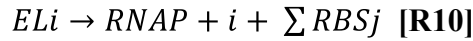


The rate of RNAP promoter clearance depends on the rate of RNAP elongation (~67 Nucleotides / Second [73, 74] and the RNAP spacing requirement (50 Nucleotides [75]). Equation E9 describes the first order reaction equation where $C_9 = 1.34 \text{ Seconds}^{-1}$.

$$Rate = C_9 * Ei \quad [E9]$$

2.11 Transcription Termination

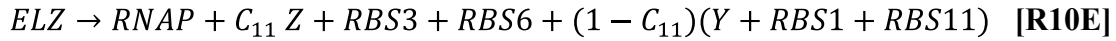
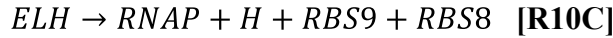
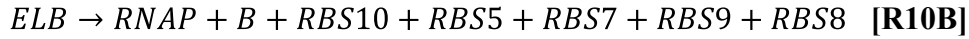
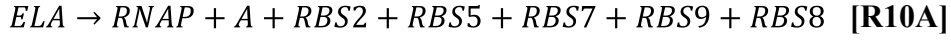
Once RNAP clears the promoter site the rate of mRNA formation is taken to be a first order reaction dependent on the length of the mRNA transcript. The products of the reaction are the reproduction of a free RNAP, a polycistronic transcript i , and species containing the ribosome binding sites j (RBS j) for each protein coded on transcript i . Transcription termination for all primary transcripts can be represented by reaction R10 where each transcript will reproduce the RNAP enzyme, create a polycistronic transcript i , and produce separate and independent ribosome binding sites (RBS) for each viral protein j coded in transcript i .



The ribosome binding sites for individual proteins on the polycistronic transcript, RBS j , are considered independent species. There are two reasons for making the ribosome binding sites independent. First, multiple proteins can be made from a single piece of mRNA and second the rates of protein initiation at each ribosome binding site are different. Therefore each RBS j is assumed to be a separate species from the mRNA molecule on which it is located.

The equations R10A-R10C describe the model species produced by transcripts initiated from the promoter sites DA, DB, and DH. R10D can be written as four separate equations because there are 4 separate rho-dependent terminators within the intergenic region. The different transcripts do not affect the coding region of P4 and there is no evidence to believe that

those 4 transcripts have different mRNA degradation rates. The model assumes the four possible transcripts are functionally equivalent. The production of transcripts Z and Y are produced in a single equation R10E and the amount of mRNAs Z and W produced depends on efficiency of the rho-dependent terminator described by the constant C_{11} .



Although mRNA A contains the coding sequence for the production of P10, see Figure 11, R10A does not include RBS10 as a product of the reaction. Experiments have observed increasing transcript A production does not result in an increase in P10 production, leading to a conclusion that little or no P10 is made from transcript A[16]. Further studies concluded P10 translation is limited on mRNA A because elongating ribosomes from P2 translation interfere with P10 initiation [19].

The conversion of R10A-E into E10A-E is based on the first order reaction of elongating RNAP. E10A-E is dependent on rate of RNAP elongation and the length of the RNAP transcript. The rate follows first order mass action kinetics as described by equation E10:

$$Rate = C_{10,i} * ELi \quad [\mathbf{E10}]$$

$$\text{Where: } C_{10,i} = \frac{RNAP \text{ Elongation Rate}}{Nucleotide \text{ Length of Transcript } i-50}$$

The rate constant $C_{10,i}$ is dependent on the rate of RNAP elongation and the length of the transcript. The nucleotide length is reduced by 50 to account for the 50 base pairs that have already been transcribed due to clearing of the promoter site.

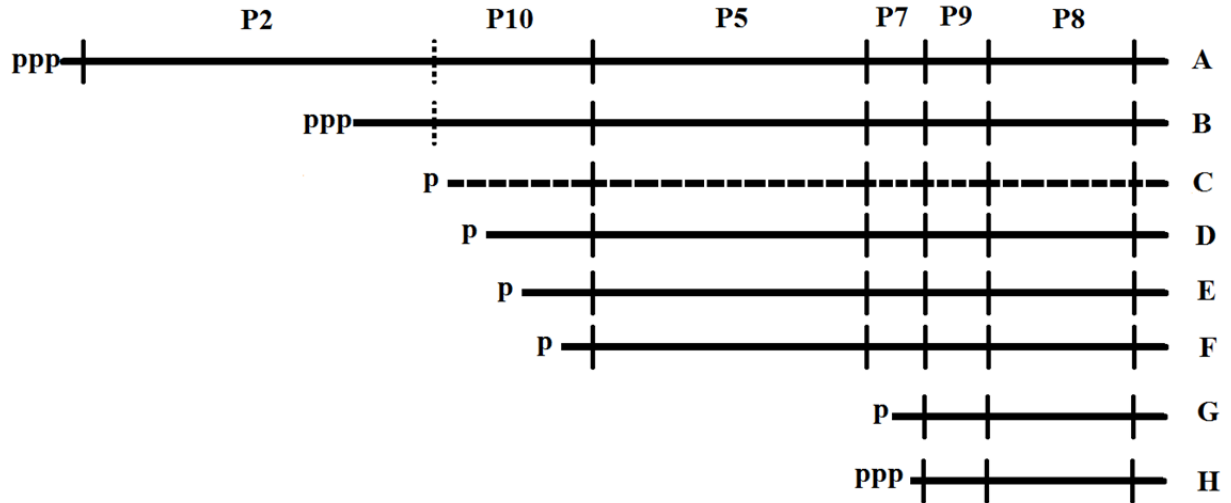


Figure 11 – A visual depiction of the mRNA strands produced directly from DNA (A, B, and H) and the transcripts produced from mRNA degradation (B, D, E, F, and G). The picture is adapted from [16]. It should be noted mRNA C is a dashed line to indicate it was left out of the model.

Table 6 – The description, symbol, value, and source for the parameters used in the transcription. Parameters with reference of N/A indicate a parameter that was fitted to observed experimental results

Parameter Description	Symbol	Value	Reference
Transcription			
RNAP Binding to Promoter Site A	$C_{8,A}$	$6E-3 \text{ (Molecule*Second)}^{-1}$	[21, 69]
RNAP Binding to Promoter Site B	$C_{8,B}$	$1E-2 \text{ (Molecule*Second)}^{-1}$	[21, 69]
RNAP Binding to Promoter Site H	$C_{8,H}$	$6.5E-2 \text{ (Molecule*Second)}^{-1}$	[21, 69]
RNAP Binding to Promoter Site W	$C_{8,W}$	$2E-3 \text{ (Molecule*Second)}^{-1}$	N/A
RNAP Binding to Promoter Site Z	$C_{8,Z}$	$2E-4 \text{ (Molecule*Second)}^{-1}$	N/A
RNAP Clearing the Promoter Site	C_9	$1.34 \text{ (Second)}^{-1}$	[73-75]
RNAP Elongation of mRNA A	$C_{10,A}$	$3.6E-2 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA B	$C_{10,B}$	$7.2E-2 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA H	$C_{10,H}$	$2.5E-1 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA W	$C_{10,W}$	$5.5E-2 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA Z	$C_{10,Z}$	$4.3E-2 \text{ (Second)}^{-1}$	[73, 74]
Efficiency of Rho-Dependent Terminator Distal to P3	C_{11}	0.60 (Unitless)	[5]

2.12 mRNA Processing and Degradation

As seen in Figure 11, additional mRNA species beyond those produced by direct transcription of the genome are present and contribute to the production of phage proteins. The processing and degradation of mRNA depends on the activities of multiple different enzymes that fall into two general categories. The first category involves exonucleases which degrade mRNA starting at the termini of the mRNA strand [76]. The second group involves endonucleases that cut at specific regions within the mRNA transcript [76]. Phage encoded mRNA are processed by both the exonucleases and endonucleases enzymes to maintain appropriate levels of mRNA production.

The processing and degradation of mRNA produced in the highly transcribed region (A, B, and H) has been extensively studied. Primary mRNA transcripts A and B have been shown to degrade into smaller and more stable mRNA species which are denoted as species C-G, see Figure 11 in previous section. At least three different pathways have been identified for processing of the primary transcripts A, B, and H into species C-G. The major pathway utilizes the endonuclease ribonuclease E (RNase E) that makes specific cleavages within the mRNA strands. The RNase E dependent mRNA processing pathway is described in reactions E11-E16 [22, 77], as shown in Table 7 and by the blue lines in Figure 12. A minor mRNA processing pathway exists which utilizes the *E. coli* ribonuclease III (RNase III) enzyme that degrades mRNA A and B into transcript C rather than D. Since mRNA C is only present at low levels and degrades very quickly, the pathway has been excluded from the model. The third mRNA transcript processing and degradation is done by exonucleases, which are assumed to be able to degrade any transcript D-H completely in one step. The assumption of one step complete physical degradation is probably unrealistic but exonucleases degradation should very quickly result in functional degradation of the mRNA species. The Approximately 30 percent of all

processed transcripts are estimated to be degraded through the exonuclease pathway [77]. The exonuclease pathway is described by equations E17-E21 in Table 7 and is represented by the red lines in Figure 12.

Table 7 – Degradation reaction equations and rate equations for the more frequently transcribed region.

Reaction Number	Reaction	Rates
E11	$A + RBS2 \rightarrow D$	$Rate = C_{12,A} * [A]$
E12	$B + RBS10 \rightarrow D$	$Rate = C_{12,B} * [B]$
E13	$D \rightarrow E$	$Rate = 0.7 C_{12,D} * [D]$
E14	$E \rightarrow F$	$Rate = 0.7 C_{12,E} * [E]$
E15	$F + RBS5 \rightarrow G$	$Rate = 0.7 C_{12,F} * [F]$
E16	$G \rightarrow H$	$Rate = 0.7 C_{12,G} * [G]$
E17	$H + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = C_{12,H} * [H]$
E18	$D + RBS5 + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,D} * [D]$
E19	$E + RBS5 + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,E} * [E]$
E20	$F + RBS5 + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,F} * [F]$
E21	$G + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,G} * [G]$

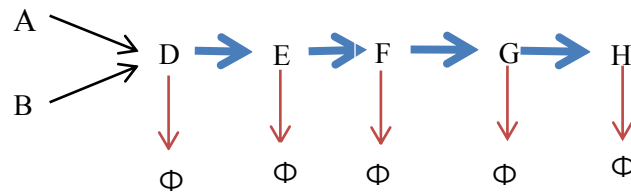


Figure 12 – mRNA transcript degradation pathway. The bold blue lines indicate RNase E dependent processing pathways and the red lines represent exonucleases degradation pathways

When mRNA A is degraded into transcript D it will have lost the translational capacity to produce P2. Therefore when one mRNA is degraded, one RBS2 must be degraded with it. The rate of degradation of A and RBS2 is only dependent on the polycistronic transcript A because

they are biologically equivalent. No ribosome binding sites are consumed in the degradation of D and E because they do not lose any translational capacity, see Figure 11.

The half-lives of all mRNA species A-H have been measured in vivo [24] and are the experimentally determined half-lives shown in Table 8. The measured half-lives were converted to rate constants assuming the overall degradation followed 1st order mass action kinetics and converted by the following equation:

$$C_{12,i} = \frac{\ln(2)}{60 * \tau_i}$$

The rate constant is divided by 60 to convert the units from minutes⁻¹ to seconds⁻¹.

Table 8 – Tabulated half-lives of mRNA species in majorly transcribed region. The table includes: the mRNA name, proteins coded for, half-life [24], and the degradation rate constant assuming 1st order mass action kinetics. *Indicates the values for half-lives which were determined to have a half-life less than two minutes and the model assumed the half-life equaled 2 minutes. # indicates a half-life was calculated using the experimentally determined ratio of mRNA B to mRNA A of 5. Therefore the half-life of mRNA A was reduced until the desired ratio was obtained.

mRNA Name	Proteins Coded	Half-Life [Minutes]
A	P2, P5,P7,P9 and P8	$\tau_A = 0.57^{\#}$
B	P10, P5,P7,P9 and P8	$\tau_B = < 2^*$
C	P5,P7,P9 and P8	$\tau_C = < 2^*$
D	P5,P7,P9 and P8	$\tau_D = < 2^*$
E	P5,P7,P9 and P8	$\tau_E = 2.5$
F	P5,P7,P9 and P8	$\tau_F = 6$
G	P9 and P8	$\tau_G = 8$
H	P9 and P8	$\tau_H = 10$

The half-lives shown in Table 8 were calculated from in vivo measurements. The measured half-lives include the effects of both the major and minor pathways. Due to limited information involving the minor pathways, only the RNase E and exonucleases pathway were accounted for in the model. The degradation pathway is modeled as depicted in Figure 12 where

the blue lines represent the RNase E dependent pathway, accounting for 70% of all degradation, and the red line represents the exonucleases pathway taken by the remaining 30% [77].

The half-lives for transcripts A-D are known to be less than 2 minutes and exact times are unknown. Transcripts B-D were given half-lives of 2 minutes in the model. Transcript A was given a half-life of less than 2 minutes by fitting the value to obtain an ratio of mRNA A to B to be of 1 to 5 as reported for the wild-type infection [16]. The estimated half-life from the model is 0.57 minutes, which correlates with an experimental estimate that it must be less than 1 minute [23].

The degradation of the mRNA from the lesser transcribed region has not been extensively studied. Therefore, a simple degradation pathway is assumed for mRNAs W, Y, and Z. The mRNAs are assumed to be directly degraded through a first order mass action kinetics, see Table 9. The model assumes that the half-life is the average between the upper and lower values of mRNA half-lives for the more frequently transcribed region because no better estimate was found in literature.

Table 9 – Degradation reaction equations and rate equations for the less frequently transcribed region.

Reaction Number	Reaction	Rates
E22	$W + RBS4 \rightarrow \emptyset$	$Rate = C_{12,W} * [W]$
E23	$Y + RBS6 + RBS3 + RBS11 + RBS1 \rightarrow \emptyset$	$Rate = C_{12,Y} * [Y]$
E24	$Z + RBS6 + RBS3 \rightarrow \emptyset$	$Rate = C_{12,Z} * [Z]$

Table 10 – The description, symbol, value, and source for the parameters used in the mRNA degradation. Parameters with reference of AVE indicate parameters which were taken as the average value between the high and low values of the measured half-lives of the more frequently transcribed region.

Parameter Description	Symbol	Value	Reference
<u>Degradation</u>			
Degradation Rate of mRNA A	$C_{12,A}$	$2.0E-2 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA B	$C_{12,B}$	$5.8E-2 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA D	$C_{12,D}$	$5.8E-2 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA E	$C_{12,E}$	$4.6E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA F	$C_{12,F}$	$1.9E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA H	$C_{12,G}$	$1.4E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA G	$C_{12,H}$	$1.2E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA W	$C_{12,W}$	$1.9E-3 \text{ (Second)}^{-1}$	AVE
Degradation Rate of mRNA Y	$C_{12,Y}$	$1.9E-3 \text{ (Second)}^{-1}$	AVE
Degradation Rate of mRNA Z	$C_{12,Z}$	$1.9E-3 \text{ (Second)}^{-1}$	AVE

2.13 Translation

In the model, each polycistronic transcript (A, B, C, etc.) is represented by a set of independent ribosome binding sites (RBSi). Translation of each gene on a single piece of mRNA is assumed to be independent and the possible interference between multiple ribosomes translating a single piece of mRNA is ignored. Similar to transcription, translation occurs in three generalized steps 1.) ribosome binding to the start codon, 2.) ribosome elongation, and 3.) ribosome termination. The number of available ribosomes is assumed to be limited, since the phage coexists with the host and only takes over about 20% of the host cell machinery [58]. The rate of translation termination is assumed to be significantly faster than the combined rate of steps 1 and 2 and is not explicitly included in the model.

2.14 Rate of Ribosome Binding to Start Codon

The ribosome binding to the mRNA is accounted for in the generic reaction rate equation R25, where RBSj is any free ribosome binding site, R is a free ribosome and the index j indicates the specific phage protein. RBSjR represents a ribosome bound to the ribosome binding site j.



Each ribosome binding site is characterized by its “strength” which is an indication of the affinity of the RBS for a ribosome and the frequency with which bound ribosomes initiate translation. Ribosome binding site strengths were calculated by using the Ribosome Binding Site Calculator developed by Voigt and Salis labs [28, 29]. The input to the calculator is the phage genome sequence. The output from the calculator is a list of ribosome binding sites and relative RBS strength. The output of the calculator for the filamentous phage genome is shown in Table 11.

Table 11– All relative promoter strengths were found using the Salis lab Ribosome Binding strength predictor unless otherwise indicated by *. P10 was predicted to have a strong ribosome binding affinity by the calculator ($RBSS_{10} = 0.92$) and was suggested to have a strong binding site experimentally [16]. However, a secondary structure exists in the mRNA which inhibits the translation of P10 decreasing the efficiency of translation initiation 10-15 fold less than P2. Therefore the relative binding site strength for P10 was taken to be 10 fold less [19].

Protein	Nucleotide Length	Relative Ribosome binding Site Strength ($RBSS_i$)
P2	1200	0.54
P10	350	0.054 *
P5	270	1.00
P7	101	1.55E-5
P9	98	0.08
P8	220	0.22
P3	1300	0.40
P6	340	0.04
P1	1040	0.08
P11	310	0.20
P4	1270	0.23

For most proteins, the rate of ribosome (R) binding to an ribosome binding site (RBS_i) follows second order mass action kinetics, described in E25A:

$$Rate = C_{13i} * RBS_i * R \quad [E25A]$$

The constant C_{13i} is a rate constant dependent on the maximum binding rate, which corresponds to the rate of ribosome binding to the P5 start codon ($C_{13,5}$) and the relative ribosome binding site strength shown in Table 11.

$$C_{13,i} = RBSS_i * C_{13,5}$$

The Salis/Voight ribosome binding site calculator was used because no experimental measurements describing the relative binding site strengths or rate constants were found. $C_{13,5}$ binding constant was chosen to be 20 (molecules*second)⁻¹ because it made the rate of translation dependent on the rate of elongation. Variations in $C_{13,5}$ of +/- 20 (Molecule Second)⁻¹ does not change the outcome of the model significantly. In other words, the rate of binding will not affect the rate of protein production, the rate of ribosome elongation will. The rate of binding was kept in the model because ribosomes are limited during M13 phage infection and the binding reaction was the simplest method of implementing the restriction.

The translation of P1, P2, P3, P5, and P10 is negatively controlled by P5 translational repression [30, 78]. The mechanism of P5 inhibition involves P5 binding to the mRNA species which codes for P1, P2, P3, P5, and P10. Data taken from Zaman et al 1990 and 1991 have concluded the translation of P2 is most strongly affected by the accumulation of P5 in the cell. Translation of the other proteins was found to be controlled in the decreasing order P2>P3>P10>P1>P5. To account for the inhibition of translation, a modified decreasing hill coefficient is included in equation E25A to make E25B:

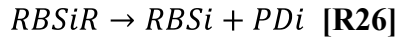
$$Rate = C_{13i} * RBSi * R * \frac{1}{1 + (P^5 / Efi * (RBSi + 1))^{n_2}} \quad \text{[E25B]}$$

E25B depends on the ratio of P5 to the RBS_i and an efficiency factor i (Efi). Increasing the value of Efi, will make protein i translation less sensitive to P5 inhibition. Decreasing the value will make translation more sensitive to P5 translation. The values for EF1, EF2, EF3, and EF10 were tuned until the experimentally determined numbers of protein molecules were produced by the model. EF2 and EF10 were tuned until approximately 1500 and 500 molecules of P2 and P10 were produced in the model [62]. P1 was assumed to be the limiting factor in phage assembly because overproduction of P1 can kill the host cell [5]. Therefore, EF1 was tuned until the number of assembly sites created were between experimentally determined number of 150-300 assembly sites [7]. The effect of P5 translation on P1 is an overestimate because we assume all of the P1 created is associated with a complete assembly site meaning the amount of P1 accumulation must be higher in wild-type phage. The efficiency factor in P3 translation was tuned until the experimental determined value of 1.5 to 1.6×10^4 molecules per cell [7] were reached. Finally, the variation of EF5 did not have a significant effect on the production of P5 but it did have an effect on phage synthesis, therefore EF5 was tuned to maximize phage production. A detailed explanation on how P5 inhibition affects the rate phage production is given in chapter 3.

The number of ribosome per cell in *E. coli* cells were measured to be between 6,800-72,000 molecules per cell [72]. Since phage infection only takes over ~20% of host cell metabolism, one can predict phage infection has a range of 1,360 – 14,400 ribosomes available for replication of phage protein species. The model assumed a static number of 7,880 which is the average of the predicted amount of ribosomes utilized by phage infection.

2.15 Ribosomes Clearing the Promoter Site

Multiple ribosomes can simultaneously be translating the same protein on the same piece of mRNA. A similar mechanism used to account for RNA polymerase spacing during transcription was employed to account for ribosome spacing during elongation. Ribosome elongation can be separated into two separate steps, the ribosome clearing the ribosome binding site and the ribosome finishing translation. The first step, described in generic reaction R26, accounts for the ribosome clearing the ribosome binding site (RBSi) which is regenerated allowing another ribosome to attach to RBSi. The protein being translated at this point is denoted as a new subspecies, PDi, which indicates an elongating ribosome that does not occupy an RBSi site.

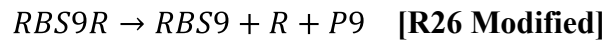


The rate of ribosome clearance is described in R26 and is expressed as a first order rate reaction in E26, where C_{14} is dependent on the rate of elongation of the ribosome (14 A.A. / S) [79] and the ribosome footprint (173 nucleotides long) [48].

$$Rate = C_{14} * RBSiR \quad [\mathbf{E26}]$$

$$\text{Where: } C_{14} = \frac{\text{Rate of Ribosome Elongation}}{\text{Ribosome Footprint}} = \frac{14 \text{ A.A./Sec} * 3 \text{ Nucleotides/A.A.}}{173 \text{ Nucleotids}} = 0.25 \text{ sec}^{-1}$$

The translation of P9 is treated differently than the other proteins because the mRNA coding for P9 is only 101 nucleotides long and is smaller than the ribosome footprint. Therefore, instead of making new species PD9 as shown in equation R22, the protein P9 will be produced instead as shown in [R26 Modified].



2.16 Translation Termination

The rate at which the ribosome completes protein synthesis is dependent on the ribosome elongation rate and the remaining number of amino acids that need to be added to the protein after the ribosome clears the ribosome binding site. The addition of the remaining amino acids follows a first order kinetics as depicted in the reaction equation R27 where the protein P_i is made and the ribosome recycled:



The rate of protein production depends on the rate of ribosome translation and the remaining nucleotides needed to be translated. The larger the protein, the more amino acids need to be incorporated into the protein and the smaller the rate at which the protein is produced. The protein production rate is reflected by the rate constant $C_{15,i}$:

$$\text{Rate} = C_{15,i} * PD_i \quad [\text{E27}]$$

$$\text{Where: } C_{15,i} = \frac{\text{Ribosome Elongation Rate}}{\text{Length of Protein Sequence} - 173}$$

P7 is not produced by an initiation of its own ribosome binding site. The binding site strength is extremely small/weak as suggested by the ribosome binding site calculator and experimental results. Experiments have suggested the translation of P7 is inhibited because a stable hairpin loop occurs around the ribosome binding site and prevents the ribosome from attaching [80]. Therefore, P7 is thought to be transcribed only through a translational coupling of P5 production. The efficiency of an elongating ribosome that has initiated translation of P5 continuing to produce P7 is 10% [80]. The model accounts for this by producing 0.1 P7 every time a single P5 is made.

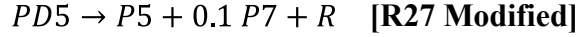
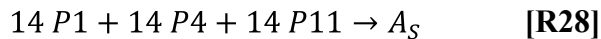


Table 12 - The description, symbol, value, and source for the parameters used in translation.# indicates the relative ribosome binding rates were known but the rate of binding was unknown.

Parameter Description	Symbol	Value	Reference
<u>Translation</u>			
Ribosome Binding to P1 Start Codon	$C_{13,1}$	$1.60 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P2 Start Codon	$C_{13,2}$	$10.8 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P3 Start Codon	$C_{13,3}$	$8.00 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P4 Start Codon	$C_{13,4}$	$4.60 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P5 Start Codon	$C_{13,5}$	$20.0 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P6 Start Codon	$C_{13,6}$	$0.80 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P8 Start Codon	$C_{13,8}$	$4.40 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P9 Start Codon	$C_{13,9}$	$1.60 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P10 Start Codon	$C_{13,10}$	$1.08 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Binding to P11 Start Codon	$C_{13,11}$	$4.00 \text{ (Molecule*Second)}^{-1}$	[28, 29] [#]
Ribosome Clearing Start Codon	C_{14}	$0.25 \text{ (Second)}^{-1}$	[48, 79]
Ribosome Elongation of P1	$C_{15,1}$	$4.8\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P2	$C_{15,2}$	$4.1\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P3	$C_{15,3}$	$3.7\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P4	$C_{15,4}$	$3.8\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P5	$C_{15,5}$	$4.3\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P6	$C_{15,6}$	$2.5\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P8	$C_{15,8}$	$8.9\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P10	$C_{15,10}$	$2.4\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P11	$C_{15,11}$	$3.1\text{E-}1 \text{ (Second)}^{-1}$	[79]
Efficiency of P5 inhibiting P1 Translation	EF1	80 (Unitless)	N/A
Efficiency of P5 inhibiting P2 Translation	EF2	$1.5\text{E-}4$ (Unitless)	N/A
Efficiency of P5 inhibiting P3 Translation	EF3	4 (Unitless)	N/A
Efficiency of P5 inhibiting P5 Translation	EF5	$1.0\text{E-}2$ (Unitless)	N/A
Efficiency of P5 inhibiting P10 Translation	EF10	$2.6\text{E-}5$ (Unitless)	N/A
Hill Coefficient for P5 Inhibiting Translation	n_2	1 (Unitless)	N/A

2.17 Assembly of New Phage

Phage assembly can be broken down into six biological steps which are outlined in Figure 7 . Step 1 involves the formation of the assembly site, A_S , which consists of 14 copies of P1, P4, and P11. The formation of the A_S is modeled by a single reaction R24:



The formation of an assembly site in one biological step is highly unlikely and R28 is a significant simplification. A more complete representation of the assembly site formation would

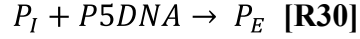
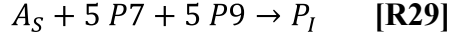
include the formation of the P4 homomultimer on the outer membrane and the interactions between P1, P11, and P4 which completes the assembly site. The details surrounding the formation of the assembly sites are unclear and R28 is currently the best way to minimize the number of determined constants and speculative reaction that would be required in a more realistic model of phage assembly site formation.

Due to the lack of information surrounding the topic, a third order rate, E24, was chosen so that if all of the proteins were available, an A_S would be made. C₁₆ was chosen so the reaction would lie heavily to the right side of the equation, favoring the formation of A_S. If C₁₆ was chosen to be too large, the numerical method used to solve the differential equations would sometimes produce negative numbers of molecules, which would indicate more assembly sites were created than the number of P1, P4 or P11 would allow. Therefore, C₁₆ was reduced until no negative numbers were observed. In the current model, assembly site formation is limited by the rate of P1 production because P1 is made in lesser quantities than P4 and P11. Since C₁₆ is chosen to be such a high rate constant essentially no free P1 exists in the system. Utilization of all free P1 molecules to form the assembly site is unrealistic but due to the lack of information it is our first approximation at modeling assembly site formation.

$$Rate = C_{16} * P1 * P4 * P11 \quad [E28]$$

The second of six steps in phage particle formation is described by R29 and involves five molecules of P7, five molecules of P9, and one assembly site to interact with each other to form an assembly site called P_I. P_I is an assembly site primed for the packaging signal of the P5DNA to bind and initiates phage elongation. Step three involves the binding of a P5DNA to P_I

produces another subspecies P_E which indicates an assembly site ready to initiate phage elongation:

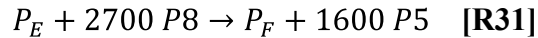


Reactions R29 and R30 follows third and second order mass action kinetics as shown in equations E26 and E27 respectively. C_{17} and C_{18} were chosen under the same conditions as C_{16} where it favors the right side of the reactions heavily:

$$Rate = C_{17} * A_S * P_7 * P_9 \quad [\mathbf{E29}]$$

$$Rate = C_{18} * P_I * P5DNA \quad [\mathbf{E30}]$$

Step 4 in the phage assembly process is phage elongation which involves the addition of 2700 P8 molecules and the shedding of the 1600 molecules of P5. Biologically this occurs in ~2700 reactions where P8 molecules replace P5 molecules. The model simplifies this process into a single step. It occurs through a one-step reaction as shown in R27 where P_F is the completed phage particle waiting to detach from the cell:



The rate equation for R27 is modeled as a first order mass action kinetics because the assumption is that the rate of phage assembly is independent of the amount of P8 molecules in the cell. In other words, the rate will be the same whether there is 10,000 molecules of P8 or 100,000 molecules of P8. As long as there are P8 molecules in the cell, phage elongation will occur.

$$Rate = C_{19} * [P_E] * \frac{P8^{n_3}}{K_{m1}^{n_3} + P8^{n_3}} \quad [E31]$$

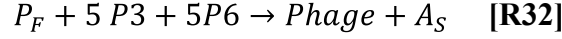
The hill coefficient prevents elongation occurring when there are less than 3000 molecules of P8. The hill coefficient, n_2 , was chosen to be 40 to assure that the model never used more P8 than available. C_{19} was calculated by using two pieces of information: First is that approximately 1000 phage are produced in 1 hour per cell [3] and secondly 225 assembly sites will be made on average [7]. Assuming all assembly sites also contain P7 and P9, it is easy to convert from $\frac{Phage}{Hour * Cell}$ to $\frac{Phage}{Second * P_I}$.

$$1000 \frac{Phage}{Hour * Cell} * \frac{1 Hour}{3600 Second} * \frac{1 Cell}{225 A_S} * \frac{1 A_S}{1 P_I} = 0.00123 \frac{Phage}{Second * P_I}$$

The determination of the rate of phage assembly is a very crude approximation because the 1000 phage/hour is an average number of phage produce in the first hour. Some cells will produce more and some less. Therefore we used a number generated from a population of cells and used it for a single cell, which may make C_{19} to be a bad estimate from true value. Additionally, the rate of phage produced in the first hour has come into question because more recent studies suggests a burst size of 250 phage particles produced in the first hour post infection [81]. If 250 phage particles are produced in the first hour then the rate of phage assembly will decrease. The effects of how decreasing the rate of phage production has on the model simulation are discussed in 3.6.

Step 5 involves P3 and P6 to bind to the end of packaged DNA to detach the assembled phage particle from the cell membrane. Without either protein, termination will not occur and another P5DNA can initiate cell elongation again [8]. Once P3 and P6 attach to the end of the phage particle, a conformational change in the P3 and P6 proteins must occur for the phage to be

released from the cell [8]. The conformational change is assumed to occur sufficiently faster than the attachment of the P3 and P6 molecules and thus is model by R28.



$$Rate = C_{20} * P_3 * P_9 \quad [\mathbf{E32}]$$

Step 6 involves the nascent phage in the extracellular space. Currently, all rate constants for phage assembly are unknown except for the rate of P8 elongation. The unknown rates are assumed to be sufficiently high to where equation E28 will be the rate determining step. All the other rate equations are necessary to evaluate if there are sufficient proteins available to initiate and terminate phage synthesis.

Table 13 - The description, symbol, value, and source for the parameters used in the phage assembly. Parameters with reference of N/A indicate a parameter that was chosen so the reaction would lie on the right hand side of the equation.

Parameter Description	Symbol	Value	Reference
<u>Assembly</u>			
Assembly Site Formation	C ₁₆	0.1 (Molecule ² *Second) ⁻¹	N/A
P _I Formation	C ₁₇	0.1 (Molecule*Second) ⁻¹	N/A
P _E Formation	C ₁₈	0.1 (Molecule*Second) ⁻¹	N/A
Phage Elongation/P8 Addition	C ₁₉	1.2E-3 (Second) ⁻¹	[7, 82]
Phage Detaching from Cell Membrane	C ₂₀	0.1 (Molecule ² *Second) ⁻¹	N/A
Number of P8 Molecules Needed for Elongation to Occur	K _{m2}	3000 (Molecules)	N/A
Hill Coefficient	n ₃	40 (Unitless)	N/A

Chapter 3: Results and Discussion

The computational model as described in the previous section was executed in Matlab with the aid of Simbiology. Simbiology is a supplemental Matlab program that translates the reaction based model into a set of appropriately formatted ordinary differential equations. The model structure was input into Simbiology. Each run of the model was executed with a Matlab m-file script that allowed the modification of all of parameters, initial concentrations of the various reaction species, the length of simulation, and execution of the simulation. The set of differential equations constituting the model were solved numerically using the built in numerical solver ODE 23t, with restriction of the maximum step size to be 5.0 seconds and a relative tolerance of 1.0×10^{-3} . The numerical solver was chosen due to the speed of simulation completion. The accuracy of the solver was not considered when choosing a numerical solver.

3.1 Validation of the Model

The model produces a series of concentration (molecules per cell) vs. time (minutes) profiles for all of the chemical species defined in the model. The following section discusses the similarities of the model results to published experimental observations accumulated during the past 50 years of intensive study of M13 biology conducted by multiple research groups from around the world. Some of the positive correlations between model and experimental observation were a direct result of modifying an experimentally undetermined parameter to achieve the positive correlation. We did not modify model parameters that were derived from experiments. Positive correlations between the modeled and experimentally observed behavior of the phage also exist that were not an intended consequence of modifying a specific parameters. In all cases we indicate where specific pieces of information derived from experiments were used to parameterize the model. We highlight instances where the model makes independent prediction

about unknown or unquantified aspects of biology. The following section describes the outcome of the model along with relevant descriptions of measured phage biology from the literature.

3.1.1 Production of M13 Phage DNA

The production of M13 phage DNA is highly complex process that involves the balanced action of three phage proteins P2, P10, and P5 as well as the enzymatic functions of *E. coli* DNA polymerase 1, DNA polymerase 3, RNA polymerase, DNA ligase, and *E. coli* single-stranded binding protein. As described in greater detail in the introductory chapter. DNA synthesis occurs in three separate phases. The initial phase involves the conversion of the entering single-stranded phage genome into a double-stranded replicative form. Early in the infection process the dsDNA replicative form replicates, producing a pool of double-stranded phage genomes in the cell. The mode of replication of the dsDNA requires the action of the phage protein P2 and proceeds through a rolling circle replication mechanism that initially produces ssDNA phage genomes that are converted to dsDNA forms by host enzymes. The third phase of DNA replication involves a switch from production of double-stranded replicative form to single-stranded DNA. The switch is a result of the action of the phage single-stranded DNA binding protein P5 that sequesters newly synthesized ssDNA molecules that are incorporated into progeny phage. The timing and extent of phage DNA production is controlled through competing protein-protein, protein-dsDNA, protein-ssDNA and protein-mRNA interactions.

We begin our description of the model's output by examining the production of double stranded replicative form DNA. Phage dsDNA synthesis depends on the competing rates of DNA polymerase III (DP3) and P5 binding to ssDNA. The phage protein P5 is a ssDNA binding protein which can prevent the conversion of ssDNA to dsDNA. At low concentrations of P5, DNA polymerase III can freely bind to ssDNA and convert it to dsDNA. At high quantities of P5, the rate of P5 binding to ssDNA to create a P5/ssDNA complex, called P5DNA, outcompetes

the rate of DP3 binding and reduces the amount of dsDNA synthesis. The accumulation of total dsDNA (sum of RF1 + RF2 + RF3/DP3 complex) is dependent on the competing rates of dsDNA synthesis (E1) and P5DNA synthesis (E7).

$$\text{Rate} = C_1 * \text{ssDNA} * \text{DP3} \quad [\text{E1}] \quad || \quad \text{Rate} = C_7 * \frac{P5^{n_1}}{k_{m1}^{n_1} + P5^{n_1}} \text{ssDNA} * P5 \quad [\text{E7}]$$

Figure 13, depicts the concentration vs time profile for the build-up of the various forms of dsDNA, as predicted by the model. The amount of dsDNA reaches a plateau at approximately 51 strands (21 molecules of RF1, 27 molecules of RF2, and 3 molecules of RF2/DP3) at 20 minutes post infection. The accumulation of dsDNA synthesis produced by the model is consistent with experimental observation of 30 – 200 strands by 30 minutes post infection [10-12]. The consistent observation between the model and experiment occurred because the rate constant C_7 , an unknown second order rate constant which directly increases or decreases the rate of P5DNA formation, was tuned until approximately 50 strands of dsDNA was accumulated in the cell. The ratio of dsDNA forms RF2 to RF1 was set to reflect the experimentally observed value of 60 percent of total dsDNA in the RF2 form. The forward and reverse rates of equations E3 describing P2 binding to RF1 DNA (constants C_3 and C_4) were chosen to produce the experimentally determined 3:2 ratio of RF2 to RF1 found in an in-vitro study [61]. A description of how the rate constants were chosen in relation to each other can be found in the “P2 Nicking RF1 DNA” section under the model description section.

$$\text{Rate} = C_3 * \text{RF1} * \text{P2} - C_4 * \text{RF2} \quad [\text{E3}]$$

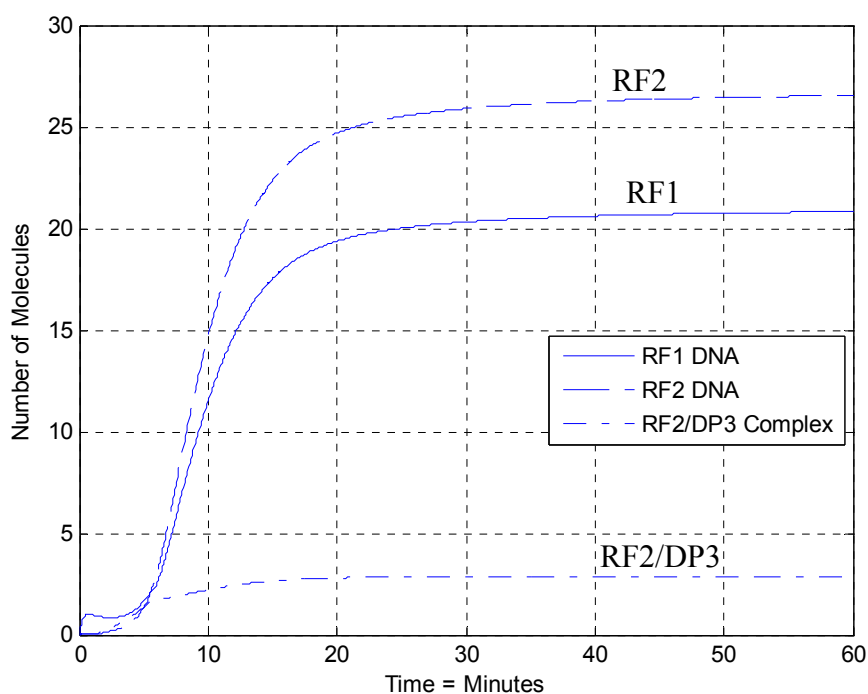


Figure 13 –Time course of phage dsDNA production. The solid line indicates un-nicked supercoiled dsDNA (RF1), the dashed line indicates relaxed dsDNA that is nicked by P2, and the dash-dot line indicates the RF2 and DNA polymerase 3 (DP3) complexes.

Figure 14 depicts the concentration of free ssDNA and ssDNA bound to a DNA polymerase 3 (DP3). According to Figure 14, a maximum of four strands of ssDNA accumulates in the cell at approximately eight minutes post infection. Early accumulation of ssDNA occurs because a maximum of one DP3 is utilized for the conversion of ssDNA to RF1, as shown by the formation of approximately one ssDNA/DP3 complex. Only one DP3 is used for dsDNA synthesis because only three DP3 enzymes are allocated to phage DNA synthesis in the model. The greater share of DP3 activity is allocated to ssDNA synthesis, acting on RF2 vs ssDNA in the form of the DP3/RF2 complex. Free ssDNA does not accumulate late in the infection because the accumulation of P5 in the cell reaches a point where the rate of P5DNA synthesis outcompetes the rate of DP3 binding to ssDNA. P5 outcompeting DP3 for ssDNA is the primary

reason why dsDNA synthesis stops in the model. The phage employs two additional control elements that affect the timing and strength of the switch from double-stranded to single-stranded DNA synthesis. One control element involves an attenuation of P2 activity by P10 and the other mechanism involves control of translational attenuation by P5, which modulates phage protein levels.

The creation of the ssDNA/DP3 complex is an indicator of dsDNA synthesis and at about 20 minutes post infection, the number of ssDNA/DP3 molecules is significantly reduced. The timing of the stoppage of dsDNA synthesis at 20 minutes post infection directly correlates with experimental observations [13]. The timing of the switch from dsDNA to ssDNA synthesis was not intentionally parameterized into the model and can be considered as a model prediction. Since the model prediction and experimental observation correlate without being parameterized to do so, this is strongly suggesting that the model is accurately recapitulating phage DNA replication.

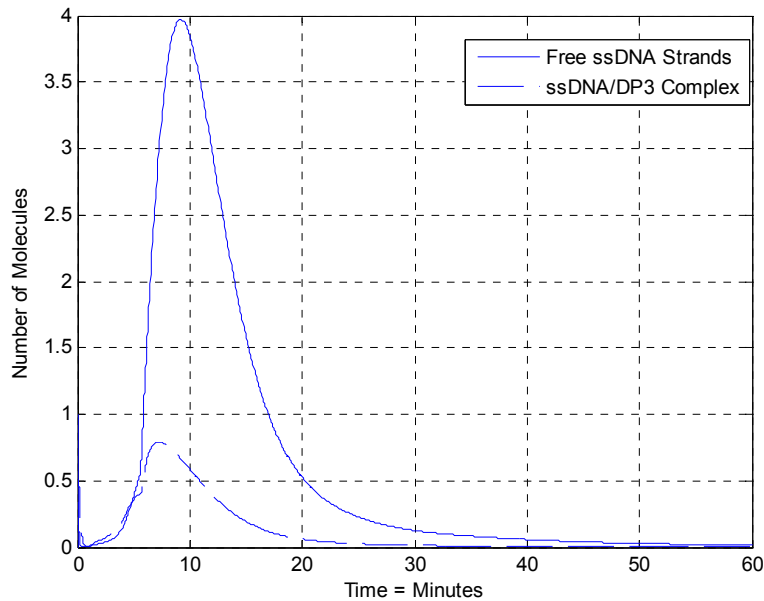


Figure 14 - The solid line indicates the number of free single-strand DNA (ssDNA) and the dotted line represents DNA polymerase III (DP3) and ssDNA complex as a function of time.

3.1.2 Production of mRNA Species

Transcription of the phage genome can be separated into two regions, the more frequently transcribed region and the less frequently transcribed region. Transcripts A, B, and H are directly transcribed in the more frequently transcribed region. Each of the transcripts A, B, and H are initiated at different promoter sites but share the same termination site. mRNA transcripts A and B are further processed and degraded into smaller and more stable transcripts D-G. One experimentally observed mRNA observed species, mRNA C, was not accounted for in the model because the degradation pathway that produces species C has not been well studied and because species C is present at very low levels in the infections (see Table 14). Three transcripts are produced in the lesser transcribed region. Two transcripts, Z and Y, arise from the same promoter but terminate at different locations. The third transcript W is part of its own cassette which only codes for one phage protein, P4. The degradation of mRNAs Z, W, and Y has not been experimentally investigated and these species are assumed to degrade in a one step process (meaning no functional intermediates are produced in the mRNA degradation process).

The ratios of mRNA A, B, and D-G predicted by the model closely correlates with the experimentally measured ratios, as seen in Table 14. mRNA C does not correlate with the experiment because it was not accounted for in the model. mRNA D, E, and F are all predicted to have a higher ratio of mRNA compared to mRNA H than the experimentally determined ratios by at least two fold. We consider a 1-2 fold difference of the model to the experimental to be a positive correlation. The variations displayed in transcripts D-F are most likely due to the very simplified model of mRNA processing and degradation. We only account for two possible pathways and in reality the mRNA processing and degradation pathways are most likely more complex and involve many pathways, as discussed in “Expanded Pathway for mRNA Degradation” section.

Table 14 - Experimental and model predicted production of mRNA normalized to the production of mRNA H. **Taken from Goodrich and Steege [77].

mRNA	Model Ratio	Experimental Ratio**
A	0.039	0.080
B	0.190	0.109
C	0.000	0.054
D	0.328	0.130
E	0.287	0.131
F	0.479	0.112
G	0.416	0.392
H	1.000	1.000

As shown in Figure 15Figure 17, all mRNA species have reached steady state or are approaching steady state by one hour post infection. The expression of mRNA B is approximately 5 times higher than the production of mRNA A, which correlates with the observed value [16] because the degradation rate of mRNA A was tuned until a the experimental ratio was obtained. We tuned the half-life of mRNA A within the bounds of the experimental measurements. The experimentally determined half-life of mRNA A is reported as an upper bound due to the limits of resolution in the experiments (≤ 2 minutes) [24]. Therefore, we believe it is acceptable to decrease the half-life until the desired ratio of mRNA A to B was obtained. Additional experimental studies have suggested the half-life of mRNA A is less than one minute which is also consistent with the value needed to produce the experimentally measured value [6]. The model predicts the half-life for mRNA A to be equal to 0.57 minutes.

mRNA D and E are shown to reach steady state at about 25 minutes post infection, which is close to the experimentally measured results of 15 minutes post infection [83]. mRNA F appears to reach steady state 50 minutes post infection and mRNA G appears to be approaching steady state at 60 minutes post infection, which is contradictory to the observed value of 25

minutes [83]. The previous statement provides further evidence that the mRNA degradation scheme may need to be revised.

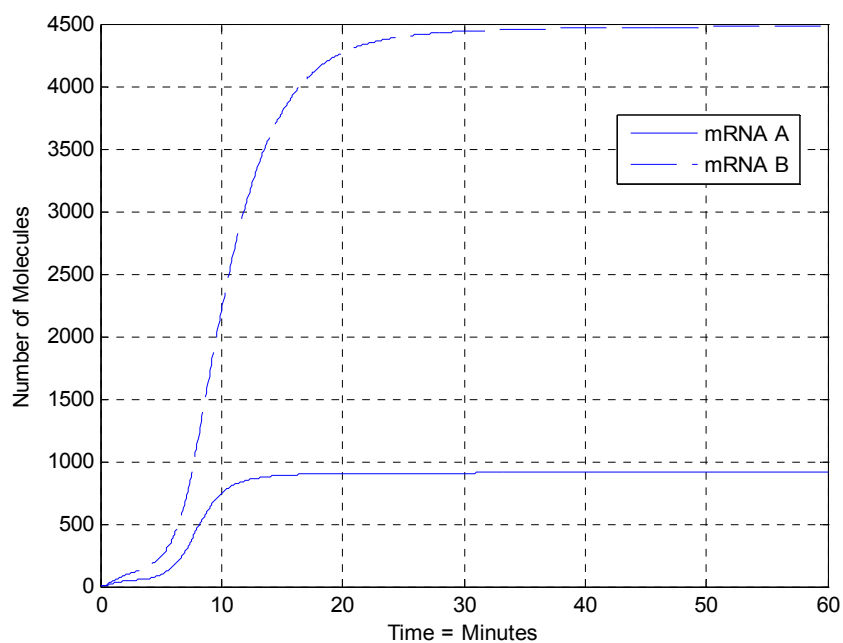


Figure 15 - amount vs. time graph of mRNA A (solid line) and B (dashed line).

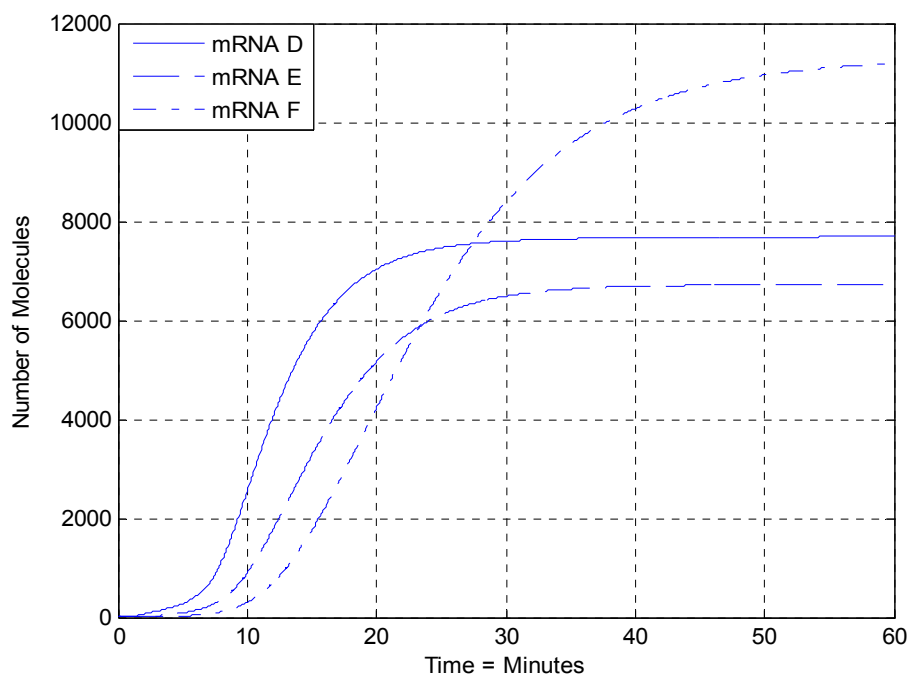


Figure 16 - amount vs. time graph of mRNA D (solid line), E (dashed line), and F(dash-dot line).

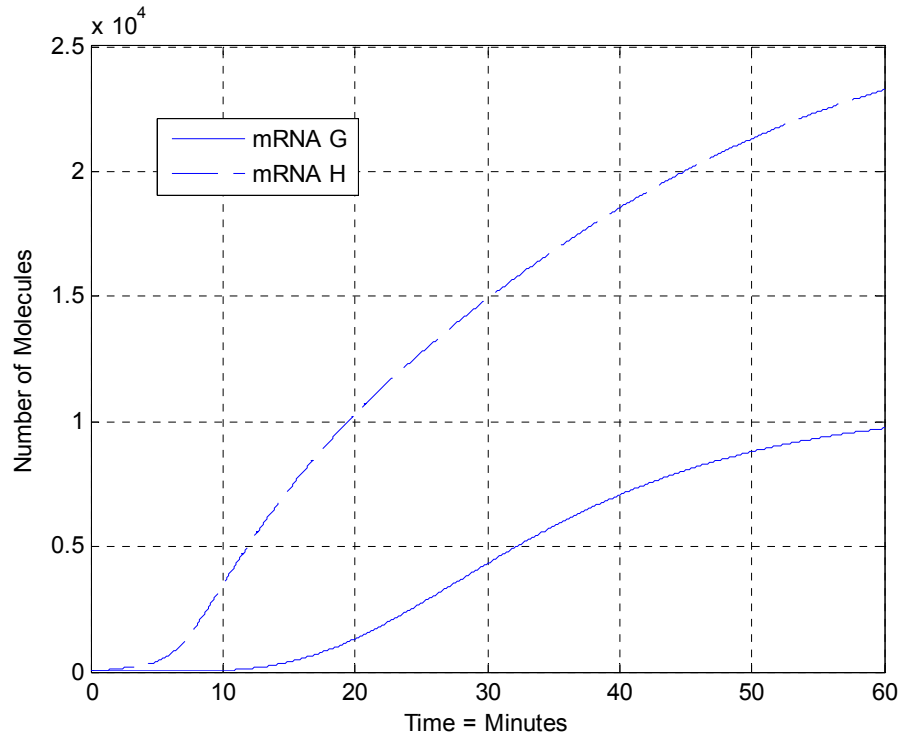


Figure 17 - Bottom, amount vs. time graph of mRNA G (solid line) and H (dashed line).

Much about the transcript formation and degradation of the lesser transcribed region is unknown. mRNA Z, Y, and W have been experimentally observed to accumulate in lower quantities than any of the mRNA species in the more frequently transcribed region [23]. The decreased accumulation of transcripts Z, Y, and W in comparison to transcripts A, B, and H is assumed to happen because transcription is initiated less frequently at promoter sites DZ and DW. The model predicts mRNAs Z and Y accumulate to lower quantities than any of the transcripts from the more frequently transcribed region, see Figure 18 Figure 19.

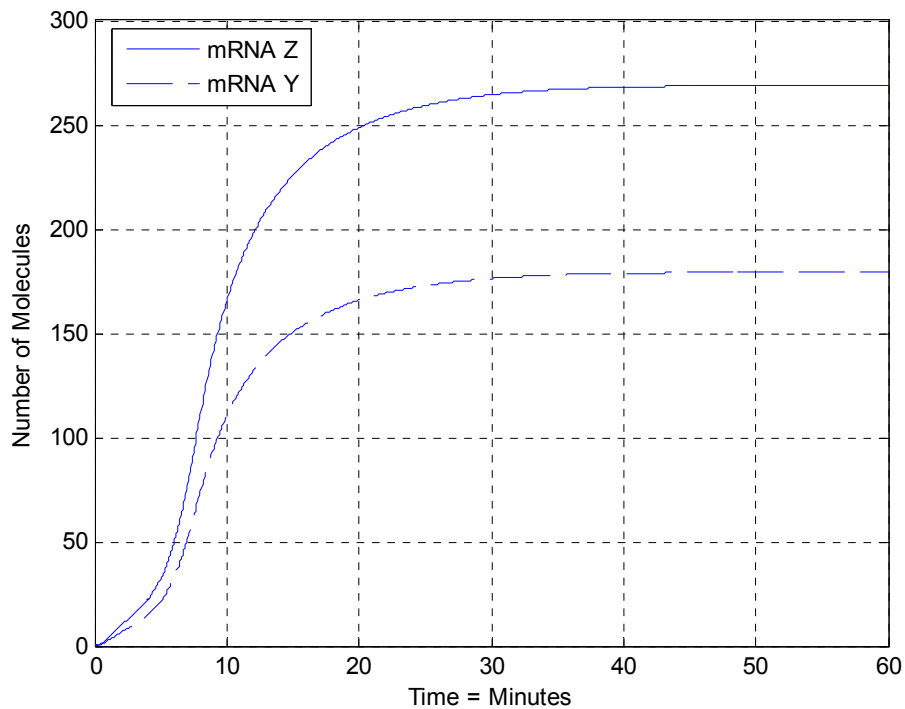


Figure 18 - The amount of mRNA Z (Solid line) and mRNA Y(dashed line) as a function of time. (W).

The model predicts that mRNA W accumulates to higher levels than mRNA A despite having a much weaker RNA polymerase III initiation rate than mRNA A. There are two possible reasons for the discrepancy. First, RNA polymerase binding to the promoter site DW is initiated too frequently. Secondly, mRNA W degrades at a higher rate than was estimated in the model. It is highly probable the degradation rate estimated in the model is drastically different from the true value because the degradation rate chosen was an average between the upper and lower bounds of the half-lives determined experimentally in the more frequently transcribed region. The half-lives for the decay of decay mRNA in the frequently transcribed region range between 8 minutes and 0.5 minutes. The high levels of mRNA W predicted in the model do not appear to have a strong effect on the outcome of the other proteins and phage produced during the model simulation.

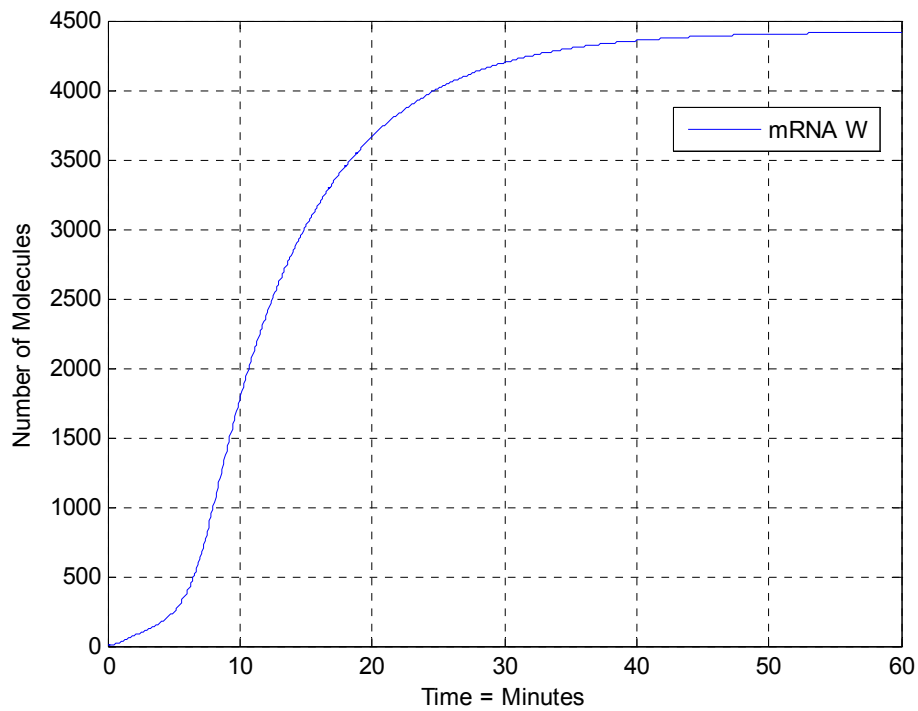


Figure 19 –The amount of mRNA W (Solid line) as a function of time.

3.1.3 Protein Production from mRNA

The rate of protein production in the model depends on the levels of mRNA, the frequency of a ribosome initiating translation of a specific protein, and the available number of ribosomes. The model assumes that the translation of viral proteins is not limited by amino acids. The model does not include protein degradation because no information is available on protein degradation rates. The accumulation of mRNA depends on the rate of transcription, and the rate of processing and degradation which was discussed in the previous section. The rate of translational initiation is different for every protein and was estimated using a thermodynamic model that calculated the binding and initiation frequency based on analysis of the ribosome binding site and secondary structure of the mRNA [28, 29]. Additionally, the translation of the viral proteins P1, P2, P3, P5, and P10 decreases as the accumulation of P5 increases.

The viral proteins can be separated into three groups which include coat, assembly, and control proteins. The coat proteins make up the protein coat of the viral particle which encapsulates the viral DNA (ssDNA). The viral coat consists of 2700 copies of P8, and five copies each of P3, P6, P7, and P9. The assembly proteins, which are P1, P4 and P11, form a channel on the cell membrane which facilitates phage assembly. Finally, the control proteins P2, P5, and P10 ensure proper accumulation of dsDNA and protein species.

3.1.4 Production of the Viral Species Involved in Regulation of Phage Infection

P5 is necessary for the formation of P5DNA (ssDNA sequestered by approximately 1600 molecules of P5) and the translational inhibition of P1, P2, P3, P5, and P10. The model produces P5 on the order of 10^5 - 10^6 molecules of P5 during the course of the infection, as shown in Figure 20. This level of P5 production is consistent with experimental observations [59]. The synthesis of P5 is rapid due to the strong ribosome binding site strength and large amount of mRNA coding for P5. The rate of accumulation of P5 reaches a steady-state because P5 inhibits its own translation. The efficiency factor 5 (EF5) in E25B was chosen to produce the optimal phage production since turning off P5 inhibition still produces P5 in the experimental range (10^5 - 10^6 molecules per cell). The mechanism through which EF5 affected phage production was not originally clear. After careful examination of the model, we have come up with a possible explanation of why P5 inhibits its own translation and why it is necessary for efficient phage infection (see the “Exploring the Effects of Removing P5 Inhibition”).

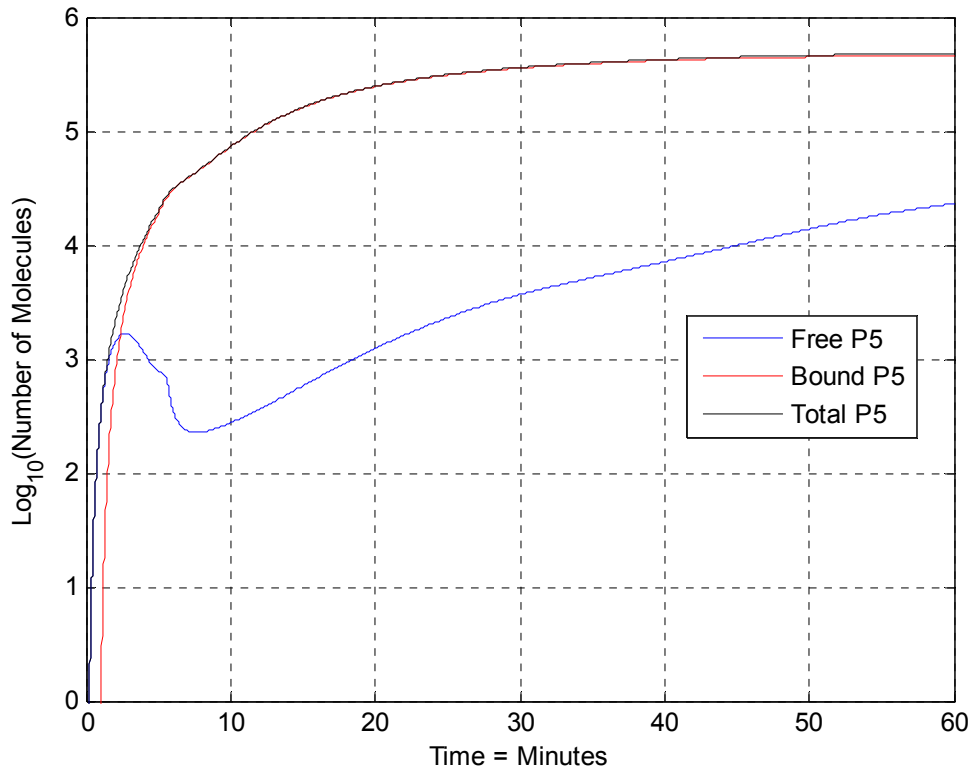


Figure 20 – the y-axis is a log based 10 graphs of all P5 species within the cell. The blue line indicates free P5 molecules, the red line indicates all P5 that is currently bound to ssDNA, and the black line indicates total P5 (bound and unbound P5 molecules)

At five minutes post infection, the accumulation of free P5 decreases and does not start to increase again until ~eight minutes post infection. The decrease in free P5 at five minutes post infection occurs because the rate of P5 binding to ssDNA outcompetes the rate of P5 translation. The amount of free P5 recovers after eight minutes post infection because the rate of recycled P5 increases from phage assembly because the number of assembling phages increases.

The model makes a prediction that at most 5 percent of P5 is free, which does not correlate with the 25 percent of free P5 molecules observed [62]. Many factors contribute to the amount of free P5 in the cell which include: the rate of P5DNA synthesis, the rate of P5 translation, the rate of phage assembly, and the binding of P5 to DNA and mRNA. If the model

is inaccurate, adjustments to any one of the three discussed factors may be contributing to the discrepancy.

Formation of P2 occurs rapidly between 1 and 8 minutes post infection, as shown in Figure 21. After 8 minutes the rate of formation rapidly decreases to zero and the concentration reaches a steady state of approximately 1800 molecules (1300 molecules of free P2 and 500 molecules of P2 bound to P10). The level of production of P2 is in close agreement with the experimentally determined level of P2 in a wild-type infection [62]. The P2 level was influenced by the EF2 in equation E25B which was tuned until the experimentally determined level of P2 was obtained. Changing the value of EF2 changes the efficiency of P5 inhibiting P2 translation. Steady state was predicted to occur earlier in the model than experimental observed value of ~20 minutes post infection [62]. Several possible reasons for the discrepancy exist. The method of P5 inhibition may not be an accurate reflection of the actual mechanism. The estimated ribosome binding site parameter may be too large thus causing an abrupt change in protein formation.

$$Rate = C_{13,2} * RBS2 * R * \frac{1}{1 + (P5 / EF2 * (RBS2 + 1))^n} \quad \text{[E25B - P2]}$$

Figure 21 also shows that no free P10 molecules are created within the system. The model prediction of no free P10 is a direct consequence of the proposed mechanism and assumptions for P2/P10 interaction. The P2/P10 association constants were set up to ensure all P10 will be converted to P2P10, which guarantees overproduction of P10 inhibits dsDNA synthesis. P2P10 represents the total number of P10 within the system. P10 production stops at about 8 minutes post infection and accumulates at approximately 500 molecules per cell, this is in close agreement with the experimental results which suggest that P10 synthesis stops at about 10 minutes post infection [62]. The steady-state level for P10 is a direct consequence of tuning the EF10 value in equation E25B.

$$Rate = C_{13,10} * RBS10 * R * \frac{1}{1+(P^5/Ef_{10}*(RBS10+1))^n} \quad [E25B - P10]$$

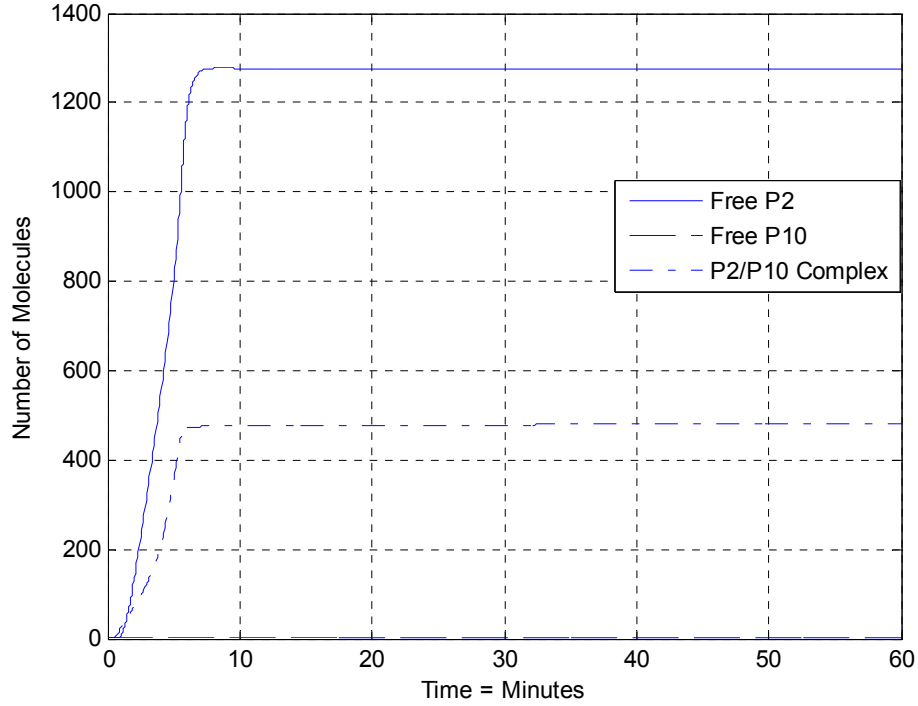


Figure 21 - shows the phage species P2(solid line), P10 (dashed line), and P2/P10 complex (dash-dot line) as phage infection progresses.

3.1.5 Production of the Viral Coat Species

Five copies of each viral protein P7 and P9 are located on the same end of the viral particle that initiates interactions with the assembly complex and is first end of the phage particle to exits out of the cell during phage assembly. The model predicts that the accumulation of phage minor coat protein P7 is $10^{4.7}$ molecules and that the accumulation of phage minor coat protein P9 is 10^6 molecules. Experimental determinations have suggested the P7 and P9 molecules must be made at low quantities because they are required in small quantities in the phage particle (5 each) and they are hard to detect [84]. However, it has been experimentally shown P7 is made on the order of 10^4 molecules [85] and the ratio of total P8 produced to the sum of P7 and P9

hold a ratio of 7:1 [84]. The model predicts the ratio to be 2.4 which is approximately 3 fold lower than the experimental value but which may be inflated due to difficulties in quantifying P7 and P9 levels. We believe the correlation between the model and the experiments is a positive indicator that we are approximating the levels of minor coat proteins well. The model and experimental results on the accumulation of P7 and P9 are consistent with each other within a factor of 3. Remember that our model is not being fit to the experimentally determined numbers but is generating protein levels through the combined action of mRNA levels, ribosome binding strengths and allocation of limited resources to production of all proteins. Since we are within a factor of 3 of the experimentally determined numbers suggest that the model is performing well.

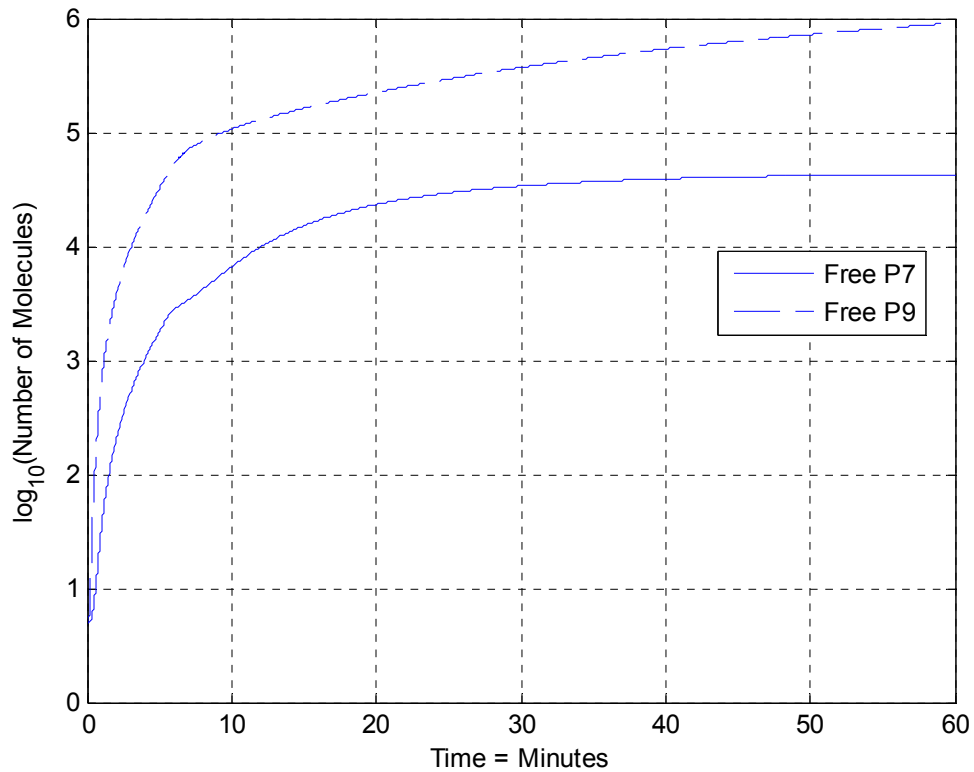


Figure 22 Time course of free P7 (solid) and free P9 (dashed) molecules. Free molecules are not associated with an assembly site or a phage particle.

In the first 8 minutes post infection, the rate of formation of both P3 and P6 appears to be equivalent, Figure 23. The rate of P6 accumulation decreases after 8 minutes post infection for

two reasons. First, P6 is beginning to be incorporated into newly formed phage particles. Secondly, P6 is being produced at a slower rate because P6 has a relatively weak ribosome binding site strength (RBSS). As a larger pool of mRNA's accumulate with stronger RBSS, those mRNA's are outcompeting RBS6 for ribosomes. P3 is predicted to accumulate until about 45 minutes post infection. P3 continues to accumulate in the cell after P6 has plateaued. P3 reaches a higher plateau concentration despite having the same number of mRNA transcripts because P3 has a much stronger RBS strength than P6. The amount of P3 accumulated is consistent with observed values of 1.5 to 1.6×10^4 molecules per cell [7]. P3 levels are controlled by the sensitivity of P5 inhibition. The parameter in the rate equation, EF3, was varied until the experimental results were obtained. The P3 and P6 protein populations decrease as the virus infection progresses because they are being exported with newly synthesized phage and the translation of both species significantly decreases, see next section on how the host resources are utilized.

$$Rate = C_{13,10} * RBS3 * R * \frac{1}{1 + (P5 / Ef3 * (RBS3 + 1))^n} \quad \text{[E25B – P3]}$$

Figure 24 shows a total of 2.3×10^6 molecules of P8 were synthesized in the cell at the end of 60 minutes. The total number of P8 molecules, labeled as “Total P8 Synthesized”, includes both P8 molecules in the cell and those packaged into phage particles. The total number of P8 synthesized is consistent with the idea that approximately 1000 phage are produced in the first hour post infection. 1000 phage particles require 2.7×10^6 molecules of P8, which is slightly more than what the model predicted. It should be noted the 1000 phage produced per hour is an acceptable value and is commented in many review articles but it is rarely cited. Recent studies have suggested phage growth may be producing ~250 phage per cell in the first generation post infection [81].

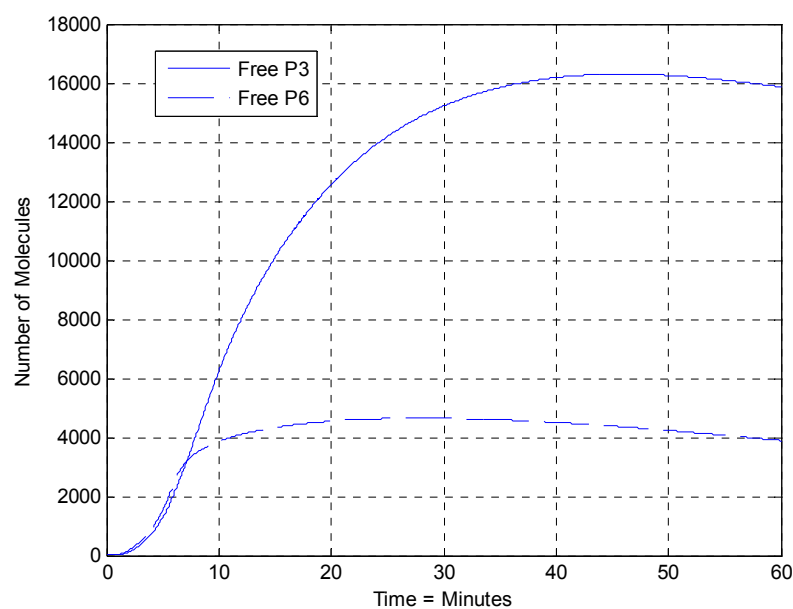


Figure 23 Concentration vs time profiles of free P3 (solid) and free P6 (dashed) molecules. Free molecules are defined as those not associated with an assembly site.

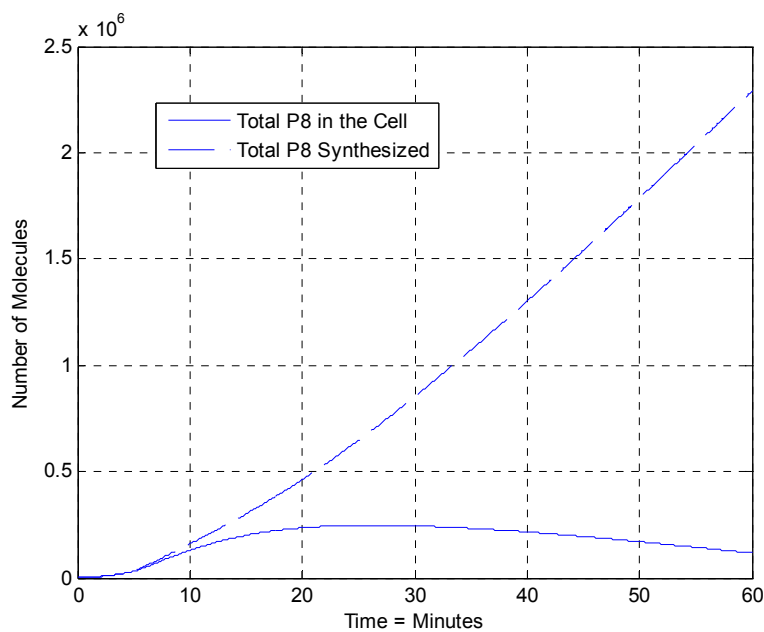


Figure 24 – Fate of P8 molecules produced during the infection process. The solid line indicates P8 that is in the cell membrane that has not been packaged into new phage. The dashed line indicates the total P8 that was synthesized in the cell which includes both P8 in the cell and P8 assembled into newly formed phage particles.

3.1.6 Production of Assembly Proteins

Approximately 3500 P1 molecules are made during the first hour of the infection process as shown in Figure 25. P1 is predicted to be the limiting protein in the formation of assembly sites because P1 translation is inhibited by P5 [30], has the lowest ribosome binding site strength compared to P4 and P11 [28, 29], and overproduction of the protein can kill the host cell [5]. Therefore since 14 molecules of P4 are assumed to be part of one assembly site and P1 is the limiting factor, then the amount of assembly sites generated is a direct consequence of the amount of P1 molecules made. The model assumes a 100 percent efficiency of the P1 molecules being incorporated into assembly sites, therefore the amount of P1 generated in the model represents a lower bound on the number of P1 molecules accumulated in the cell. Therefore, EF1 in equation E25B was varied until the range of assembly sites were between 150 – 300 total sites [7]. Experimental observations indicate that the ratio of P11 to P1 is 2 to 1 [86]. The model determined ratio of 2.8:1 closely matches the experimental ratio.

$$Rate = C_{13,1} * RBS1 * R * \frac{1}{1 + (P^5 / Ef1 * (RBS1 + 1))^n} \quad \text{[E25B – P1]}$$

We have not found any published account discussing the accumulation or steady state level of P4 within the cell. Overexpression of P4 appears to have minimal effect on the growth and viability of the cell. Transcription of P4 is controlled by its own promoter and termination site and is the only phage protein which is translated from its own mRNA strand. P4 protein levels can be manipulated within the model by changes in promoter strength, mRNA degradation rate, or ribosome binding strength. All three variables controlling P4 levels included in the model were not derived from well quantitative experimental evidence. Since there is a fair amount of uncertainty in several parameters controlling P4 levels we expect that the predicated level of expression for P4 may be less accurate than the model prediction of other proteins.

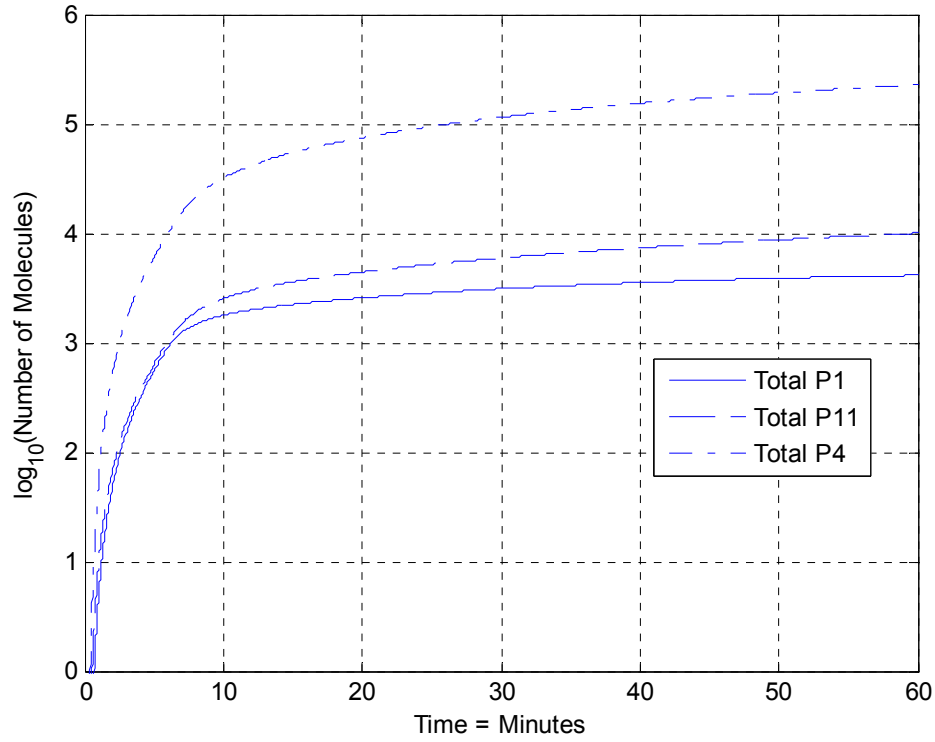


Figure 25 – Log scale plot of time course of assembly complex proteins P1, P4, and P11. The y axis is a log based 10 of the number of molecules. Each species represents the total amount of protein within the cell. This includes both free protein and protein associated with the assembly complex.

3.1.7 Production of P5DNA

The total number of assembly sites, as shown in Figure 26, is in agreement with the experimental observation of 150 – 300 total sites [7]. The agreement between experimental and model occurs because of the regulation on translation of P1 by P5 was tuned until the experimental number of P1 molecules/assembly sites was produced. P5 inhibition of the translation of P1 was controlled by the efficiency factor EF1. Increasing EF1 in equation E25B-P1 will decrease the effect of P5 inhibition on translation of P1 while decreasing EF1 increases the effect of P5 inhibition on the translation of P1. EF1 has not been determined experimentally

and it was varied until the number of assembly sites produced in the model fell in the experimentally determined range.

$$Rate = C_{13,1} * RBS1 * R * \frac{1}{1 + (P^5 / Ef1 * (RBS1 + 1))^{n_2}} \quad \text{[E25B-P1]}$$

The model predicts (which indicates it was not parameterized to match experimental results) that there are more assembly sites than available phage genomes ready to be packaged (P5DNA) during the first 30 minutes post infection. The result suggests the formation of P5DNA is limiting during the first 30 minutes of the infection. Between 30 and 50 minutes post infection, more P5DNA exists than assembly sites, but the difference is minimal. Overall, the model predicts the number of assembly sites is created in greater than or approximately equal amounts to P5DNA. Therefore the model suggests phage assembly will not be limited by assembly site formation. It should be noted, that the implantation of assembly site formation in the model was very crudely set up and it may be possible we are over-estimating the rate or the final number of assembly sites in the model.

Figure 26 also shows the number of P5DNA molecules that accumulate in the cell during the infection. The model predicts approximately 275 molecules of P5DNA accumulate in the cell by 60 minutes post infection. An in-vivo study has tested the number of P5DNA molecules in the cell by multiple different molecular biology techniques and has predicted the P5DNA molecule range between 150-350 strands of P5DNA [10]. The model prediction of 275 molecules accumulated at 60 minutes post infection correlates well with the experimental observations. No parameters were adjusted to directly achieve the correlation between the model predicted and experimental P5DNA accumulation.

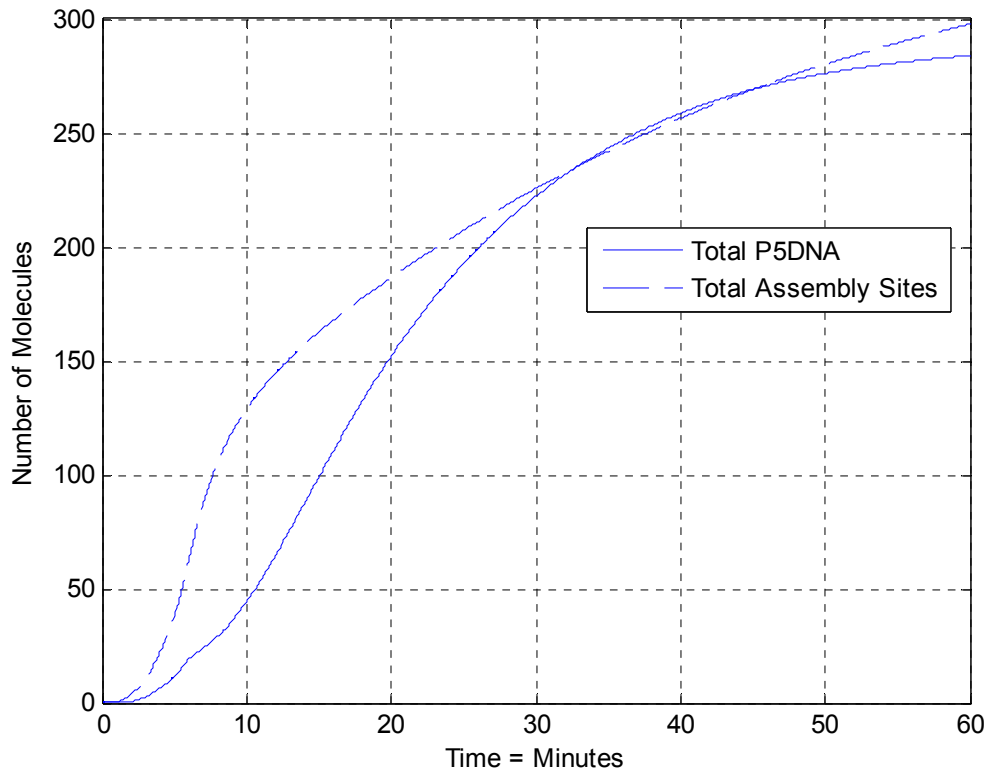


Figure 26 - The solid line indicates the total (free and bound to assembly sites) P5DNA within the cell. The dashed line indicates the total number of assembly sites within the cell.

3.1.8 Production of Phage

The rate of phage production is predicted to increase until it appears to reach a plateau around 21 phage/minutes, see Figure 28. The model produces 800 phage in the first hour after infection and correlates strongly with the approximate measured value of ~1000 phage produced [3]. The approximate value of 1000 phage produced in the first hour is in question because it was pulled from a review article which cited no source and several recent detailed studies of phage production rates suggests 250 particles of phage are produced in the first hour [81]. The model parameter relating to the rate of phage assembly was determined based on the number of 1000 phage produced by 300 assembly sites and may need to be reduced to describe case where 250 phage are produced by a single cell. Further discussion about the over-estimate on phage

elongation is discussed in section 3.6. The first phage particle is predicted to exit the cell at approximately 5 minutes post infection, which is in close agreement to the observed value of ~ 6 minutes [81].

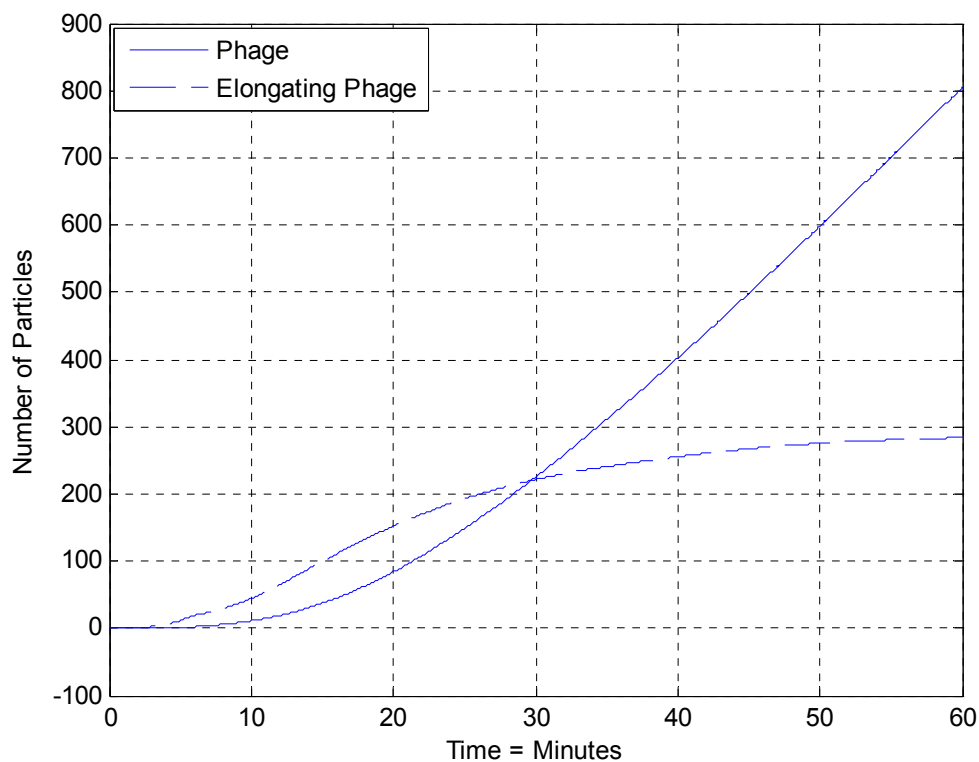


Figure 27 – Displays the rate of phage produces vs time after infection. Phage species indicates fully formed phage that have detached from the cell membrane. It also displays the number phage species which are actively being produced that have not detached from the cell membrane. Elongating phage indicate phage that are in the process of being formed.

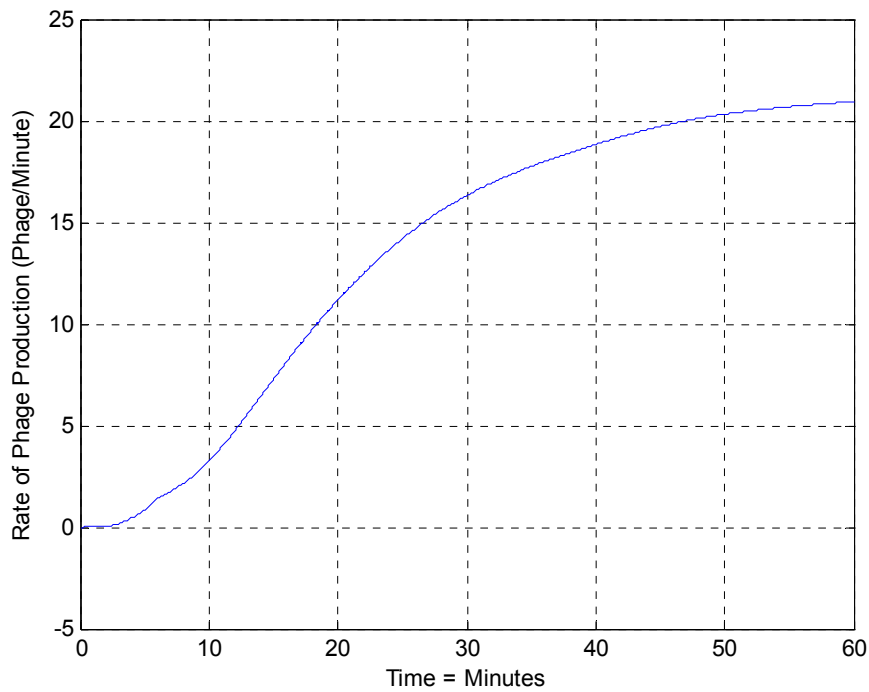


Figure 28 - The rate of phage production (Phage/minute) as a function of time.

3.2 Utilization of Host Resources

The rate of transcription and translation of phage proteins is limited by the available host RNA Polymerases III (RNAP) enzyme and host ribosomes. The amount of RNAP and ribosomes are limited by the number the host synthesizes and by the balance of resources allocated between the host and infecting phage. Based on the effects that phage infection has on cell growth, the phage is estimated to utilize 20 percent of the host cell resources [58]. Based on the composition of *E. coli* replicating with a doubling time of 1 hour, 20% of cellular resources limits the number of available RNAP to approximately 1280 enzymes and the available number of ribosomes to 7880 ribosomes. According to the model, transcription becomes limited by the available number of RNAP at approximately 15 minutes post infection and translation becomes limited by the available number of ribosomes at approximately 6 minutes post infection, see Figure 29. The

following section is concerned with analyzing how the limited resources are used during the first 60 minutes post infection.

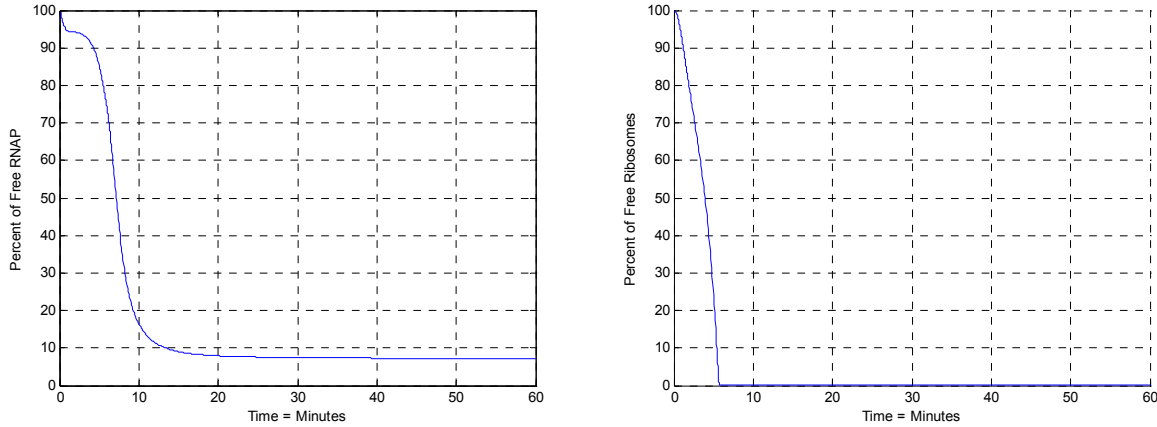


Figure 29 – (Left) The percent of free RNA Polymerase III (RNAP) enzymes not actively transcribing or bound to DNA. (Right) The percentage of ribosomes which are not bound to mRNA.

According to Figure 29, phage infection maximally utilizes 92% of the allocated RNAP enzymes available for phage infection (or 8 % of the allocated RNAP are free). Total utilization of all allocated RNAP enzymes does not occur because the number of available promoter sites/dsDNA molecules reaches a plateau at around 20 minutes post infection, Figure 13. Increasing the amount of dsDNA synthesized in the host cell will decrease the amount of free RNAP. Additionally, the free RNAP can be decreased by increasing the binding constant for RNAP binding to a promoter site.

3.2.1 RNA Polymerase Utilization

The model predicts that that approximately 40 percent of all RNAP will be devoted to synthesizing mRNA A, despite mRNA A having a weaker promoter than either mRNA B or mRNA H, see Figure 30. mRNA A controls more resources because it is the largest transcript and takes the longest to transcribe. Figure 31 shows there is a higher percentage of free promoter sites for DA than the other promoter sites DB and DH. The length of the transcript plays a role in

the control of cellular resources. An additional example where the length of the transcript plays a role in the division of cellular resources is transcription of mRNA W. Figure 32 shows 13 % of RNAP is predicted to be devoted to transcribing mRNA W, which is a larger percentage of RNAP than is involved in the transcription of mRNA H. Figure 33 shows approximately 88 percent of promoter site DW is free compared to approximately 59 percent of promoter site DH free. mRNA W transcription is initiated less frequently than mRNA H but more resources are devoted to transcribing mRNA W.

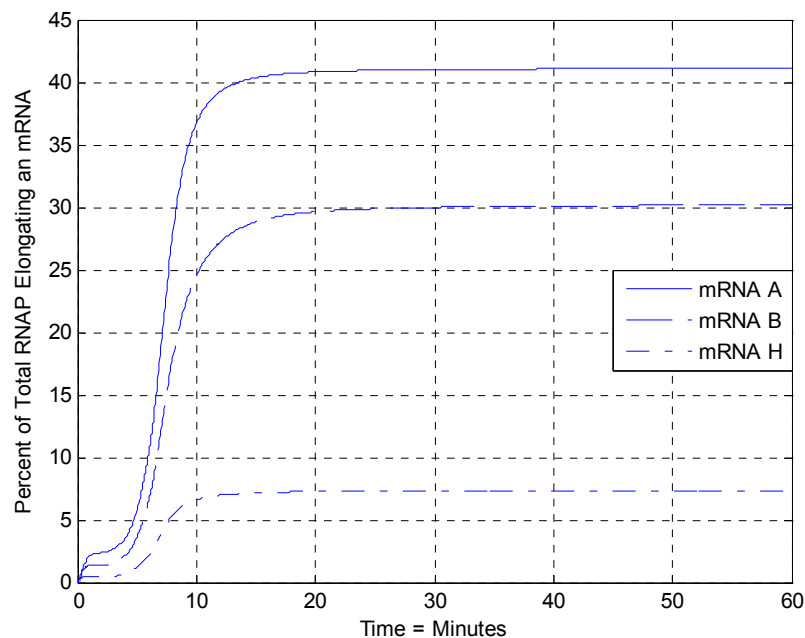


Figure 30 – Time course of RNAP elongating mRNA transcripts in the more frequently transcribed region. Percent of phage allocated RNA polymerase III (RNAP) enzymes actively translating mRNA A, B, and H.

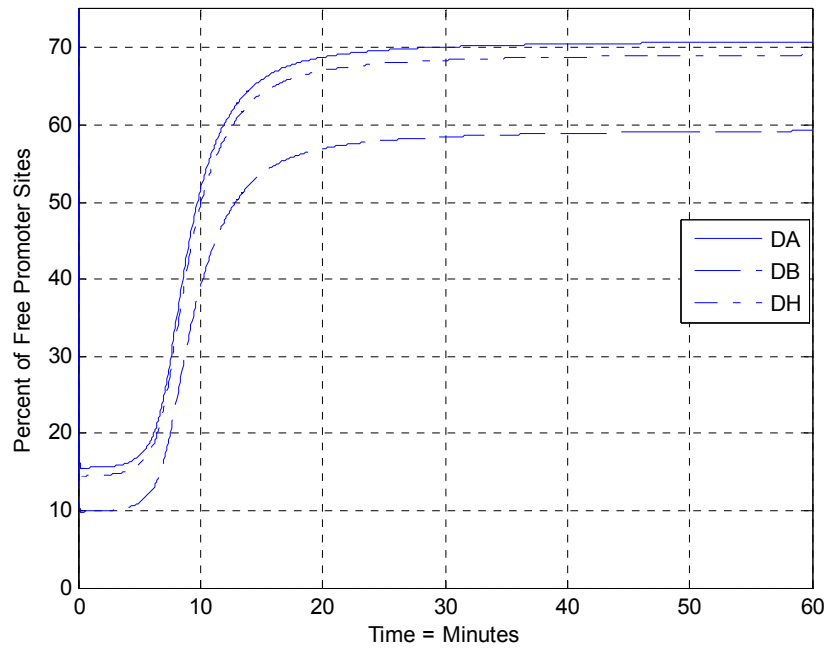


Figure 31 – Time course of RNAP occupying promoters in the more frequently transcribed region. Displays the percent of free promoter sites as a function of time for promoters DA, DB, and DH.

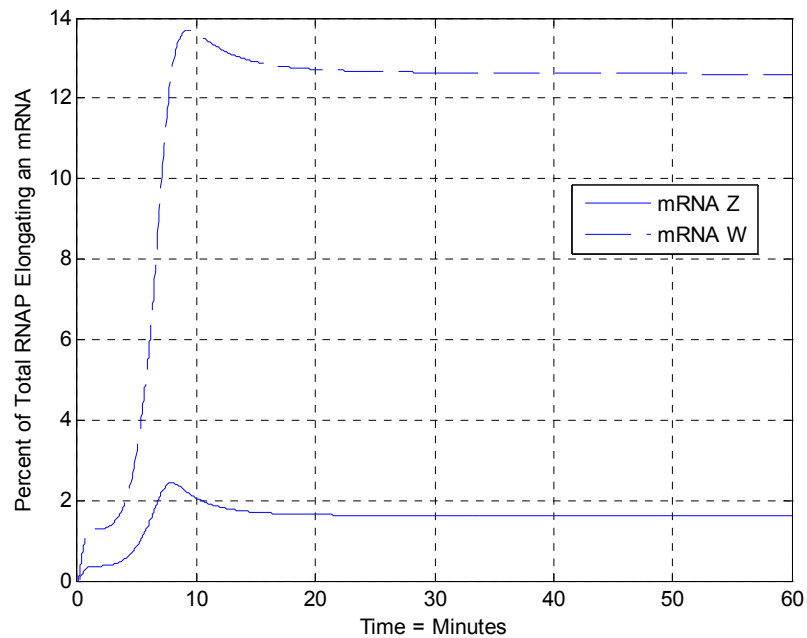


Figure 32 – Time course of RNAP elongating mRNA transcripts in the lesser transcribed region. Percent of phage allocated RNA polymerase III (RNAP) enzymes actively translating mRNAs Z and W.

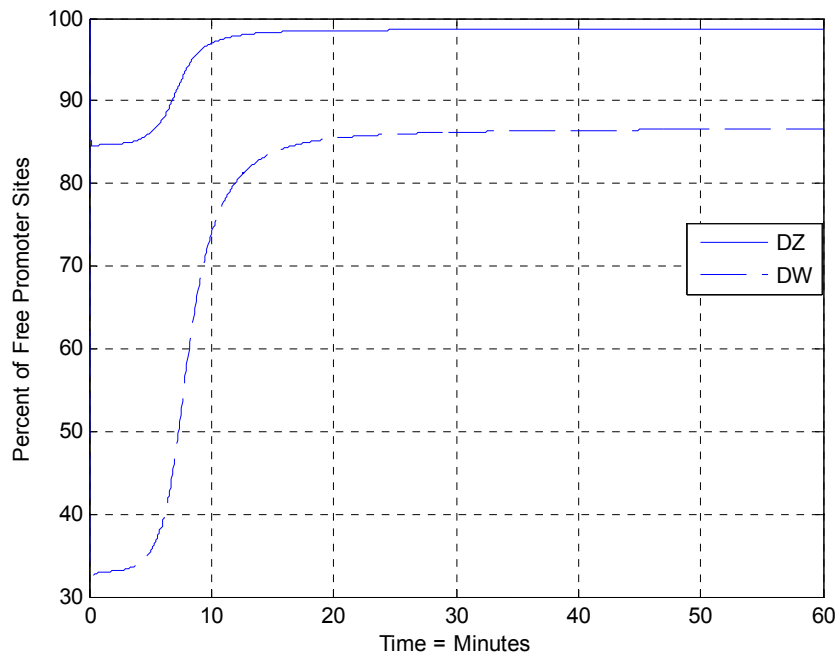


Figure 33 - Time course of RNAP occupying promoters in the lesser transcribed region. Displays the percent of free promoter sites as a function of time for promoters DZ and DW.

In the first 6 minutes post infection, the number of free promoter sites remains constant in both regions of transcription, see Figure 31 and Figure 33. At this time, the rate of formation of mRNA is limited by the rate of elongation. After 8 minutes post infection, host resources begin to become limiting and more promoter sites remain free. As expected, the weakest promoter site (Promoter A), has the greatest percentage of free binding sites. The strongest promoter (Promoter B) has the least number of free promoter sites. The model predicts at least 58 percent of all promoter sites are unoccupied. Approximately 2% of DZ promoters are occupied see Figure 33. The low level of transcription from the DZ promoter correlates with in-vitro studies that fail to see mRNA produced from DZ promoter [27].

A maximum of 70 RNAP enzymes are predicted to be elongating a single strand of dsDNA, as shown in Figure 34. The amount of RNAP/dsDNA decreases until it reaches a

plateau around 26 RNAP per dsDNA. As discussed in the model description section, RNAP needs approximately 50 base pairs between each elongating RNAP. Assuming the average length of the phage genome is 6400 base pairs, theoretically 128 RNAP / dsDNA strand can occur. The predicted maximum is half the theoretical value. No information has been found to either support or refute this prediction.

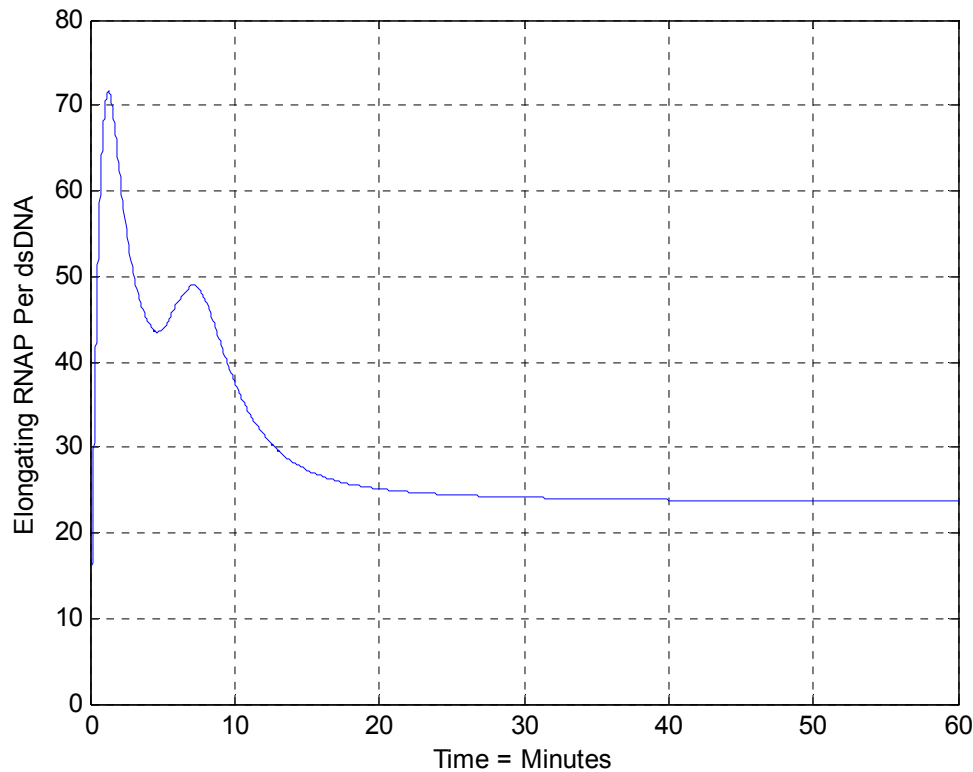


Figure 34 -The y-axis represents the number of actively translating ribosome creating each mRNA per dsDNA. In other words, the graph represents an average number of RNAP attached to a single piece of dsDNA in the actively transcribed region.

3.2.2 Ribosome Utilization

Translation of phage proteins P1, P2, P3, P5, and P10 are all negatively affected by the accumulation of P5 molecules. How the host resources are utilized strongly depends on the accumulation of P5. In other words, how the ribosomes are utilized when the P5 molecule is low will be different than the ribosome utilization when the P5 molecule concentration is high.

The model predicts a maximum of 3.5 and 0.25 percent of all phage allocated ribosomes will be translating P2 and P10 respectively, Figure 35. The model predicts a sharp decrease to zero percent of ribosomes translating P2 and P10 at approximately 8 minutes post infection. Zero percent of actively translated ribosomes means the translation of P2 and P10 stops. The fast transition between ribosomes actively translating the proteins to zero actively translating the proteins most likely does not occur in the cell. The predicted model equilibrium of P2 and P10 occurs because the translation of both proteins stops. In the real cell, the inhibitory effect of P5 on mRNA is more stochastic in nature and some P2 and P10 proteins would be produced. Total inhibition of P1, P3, and P6 occur much more gradually than P2 and P10, Figure 36. A maximum of 25 percent of the total ribosomes become dedicated to P5 translation and at 60 minutes post infection approximately 2.5 percent of ribosomes were translating P5.

Although part of the infrequently transcribed region, P4 translation takes 23 percent of the total ribosomes available for the phage infection, Figure 37. My personal speculation is that this number is unrealistically high and is probably a direct consequence of making too much mRNA W. The translation of proteins P6 and P11 reach a maximum at approximately 8 minutes post infection and decrease to a steady-state value of 0.1 and 0.1 percent respectively, Figure 38. The translation of P6 and P11 has a similar shape as the P5 inhibited proteins but the decrease in translation of P6 and P11 is caused by different mechanism. Both P6 and P11 have weak ribosome binding site strengths in comparison to other proteins. Until eight minutes post infection, the number of ribosomes is not the limiting factor determining protein translation rather insufficient mRNA production is the limiting factor. When excess ribosomes are present after all of the stronger ribosome binding sites are occupied, ribosomes will translate P6 and P11. Beyond 8 minutes post infection the rate of translation is limited by the number of ribosomes

allocated to phage infection and free ribosome binding sites of all proteins begin to accumulate. The stronger ribosome binding sites will recruit the ribosomes more frequently than the weaker ones if the ribosome has a “choice”.

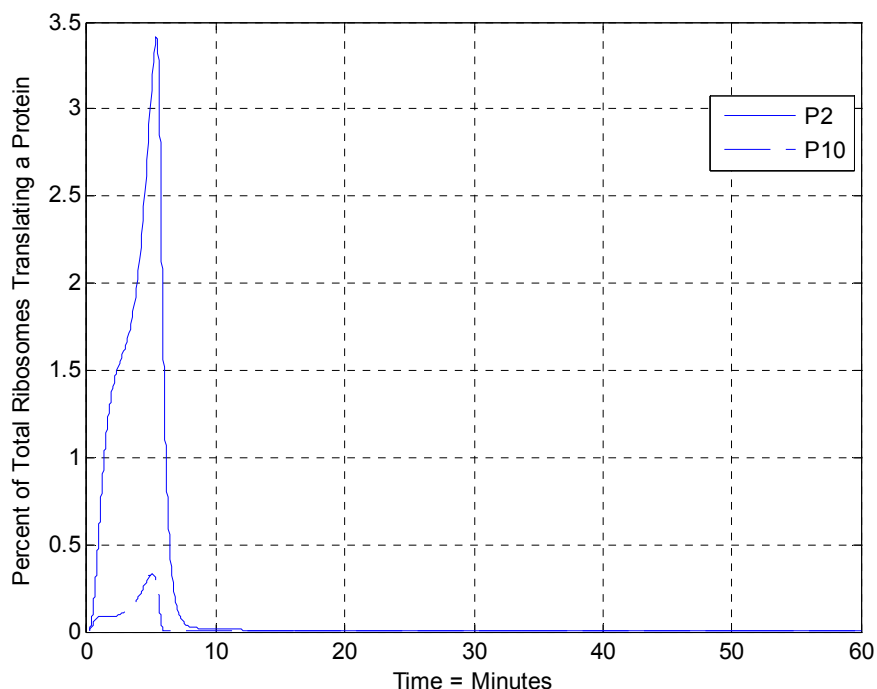


Figure 35 – The percent of allocated phage ribosomes actively translating a viral protein. The figure displays the percent of ribosomes translating proteins P2 and P10, which are both negatively inhibited by P5 accumulation in the cell.

At 60 minutes post infection, approximately 55 percent of all ribosomes are translating P8 proteins, Figure 37. P8 must be produced in the greatest quantities (10^6) because 2700 copies are needed for one phage particle, therefore it is reasonable and expected for a significant amount of available ribosomes to be dedicated to translating P8. It should be noted the translation of P8 is the only viral protein in which we observe an increase in viral translation rate as the infection proceeds and we believe it is an important event in the viral infection cycle, which is described in the next section.

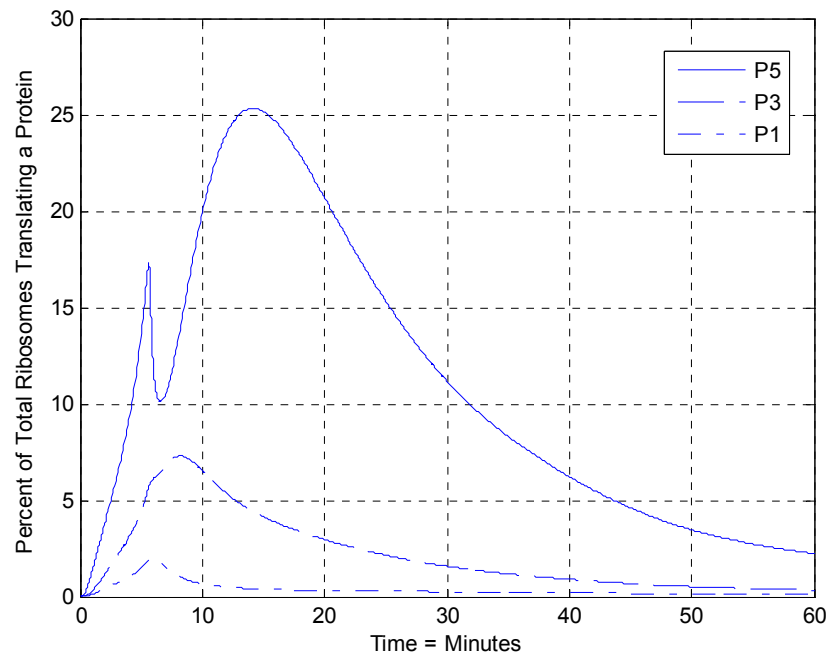


Figure 36 – The percent of allocated phage ribosomes actively translating a viral protein. The figure displays the percent of ribosomes translating proteins P5, P3, and P1, which are all negatively inhibited by P5 accumulation in the cell.

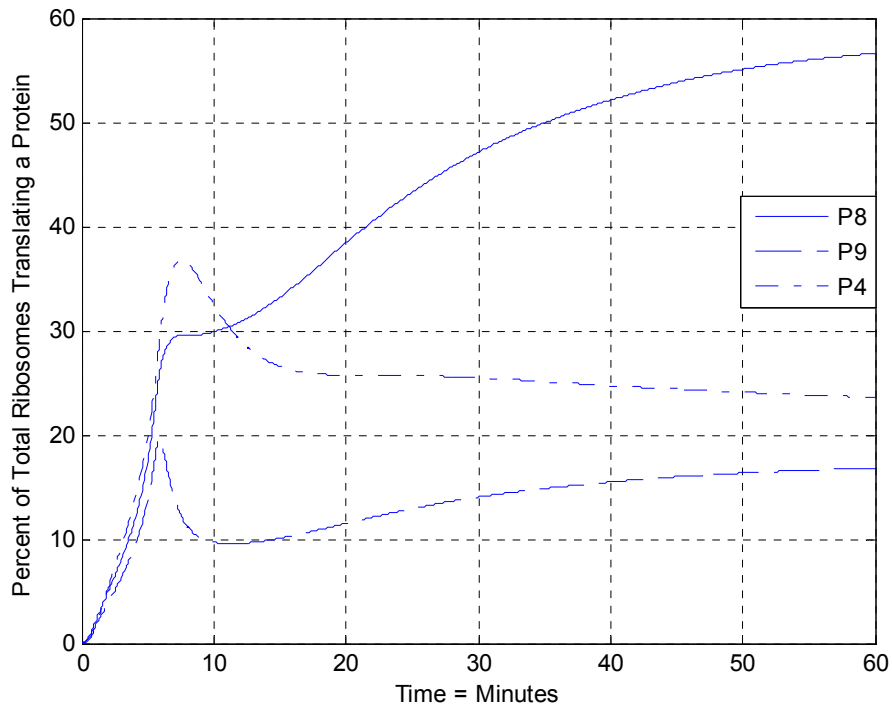


Figure 37 - The percent of allocated phage ribosomes actively translating a viral protein. The figure displays the percent of ribosomes translating proteins P8, P9, and P4, which are not inhibited by P5 accumulation in the cell.

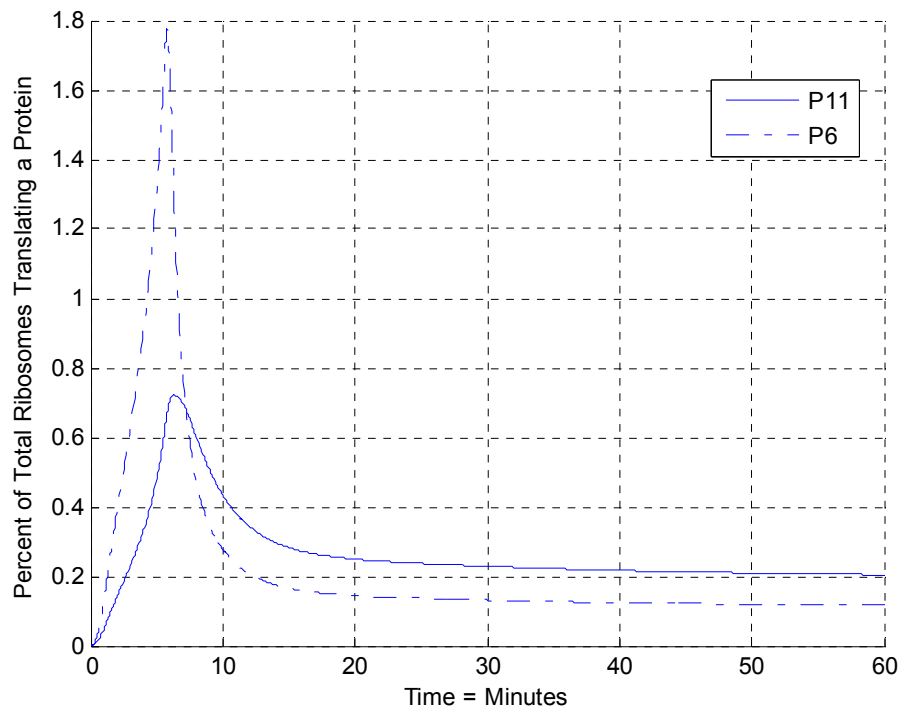


Figure 38 - The percent of allocated phage ribosomes actively translating a viral protein. The figure displays the percent of ribosomes translating proteins P6 and P11, which are not inhibited by P5 accumulation in the cell.

3.3 Exploring the Effects of Removing P5 Inhibition of Proteins 1, 2, 3, 5, 10

The following section explores the effect of fully or partially removing the P5 mediated translational inhibition of phage proteins P1, P2, P3, P5, and P10 has on phage reproduction. Fully removing the P5 inhibition will allow all proteins to be translated unrestricted by the level of P5 accumulation. Partially removing P5 inhibition will allow one phage inhibited protein to be translated unrestricted by the level of P5 accumulation. Some experimental work has been done examining the extent of P5 translational repression in plasmid systems, outside of the context of a full phage infection. This section compares the experimental plasmid system results and the computational model results directly. Since the levels of P5 in the plasmid system were not quantified, it is unclear on whether or not the results can be compared. Additionally our interpretation of the model results indicates that for some phage mechanisms of attenuation is an

integral part of the phage infection cycle. Our model shows that the attenuation attributed to P5 is sufficient to explain the timing and control of levels in the switch from dsDNA to ssDNA synthesis and further that the mechanism of P5 inhibition can explain the carrier state of the infected cell. Our model makes assumptions about the cellular resources available to phage production, particularly that the phage has infinite metabolic resources and energy. Some of the attenuation we attribute to P5 could be the result of metabolic limitations that we have not included in the model.

Experiments performed in the laboratories of Model and Konings have identified and examined the effects of translational attenuation in the phage life cycle mediated by phage protein P5 directly interacting with mRNA and preventing translation. The most complete and quantitative study of the effect of P5 inhibition of translation were performed in the Koning laboratory [30, 31]. Zaman and Konings were able to estimate the translational attenuation caused by P5 on each of the phage proteins P1, P2, P3, P5, and P10. Zaman and Konings generated a series of plasmids which contained the upstream promoter regions of the various phage proteins that show translation repression (one of the P5 inhibited proteins) and fused it to the β -galactosidase gene. The general experimental strategy was to infect a cell with two plasmids. One contains gene 5 under the arabinose inducible promoter. The second plasmid contains one of the P5 inhibited proteins translation initial signal which is fused with the lacZ gene. The production of the P5 inhibited protein is directly correlated with β -galactosidase activity. They measured the β -galactosidase activity in the presence and absence of arabinose, which can be interpreted as the presence and absence of P5. Once they measured the activities, they took the ratio of β -galactosidase activity in the presence and absence of P5 and those are the numbers expressed in the last row in Table 15. In summary, they tested the translation of each

protein (the ones inhibited by P5 translation) in individually but parallel experiments and came up with a ratio of expression in the absence and presence of P5. It should be clear that the experiment did not replicate phage but only two phage proteins were made in the cell at a time because the phage genome was not inserted into the cell.

The model can replicate the experiments by turning P5 inhibition off of individual proteins by changing the value of the efficiency factor (Efi) to 10^{30} . Changing the value to a large value ensures the total amount of P5 produced will not have an effect on translation. The general model experiment consisted of running the model with all P5 inhibition on and recording the final number of viral proteins accumulated in the cell at the end of 60 minutes. Then the model was run again with the P5 inhibition on the translation of a single protein turned off. Again the final number of viral proteins accumulated in the cell at the end of 60 minutes was recorded. The ratio was calculated by taking the value without inhibition and dividing it by the value with inhibition. The value for the efficiency factor changed to 10^{30} was reset back to the originally parameterized value and the next efficiency factor was changed to 10^{30} . Once all of the individual experiments were run, a final model experiment was run where P5 inhibition was turned off for all proteins.

There are two very important differences between the model and lab experiments. First, the model turns off P5 inhibition by making the promoter site immune to P5 accumulation. The lab experiment controlled P5 expression to measure the difference in translation with and without P5 translation. The second difference is we are removing P5 inhibition and its effects on infection process as a whole. The lab experiments did not infect their cells with a complete phage genome. Although we see positive correlations between the model and experiments, the implications of P5 inhibition on the overall phage life cycle are not completely clear. Table 15

shows that there are close matches between the model predictions and the experimental measurements for all proteins except for P10.

Table 15 – Compilation of protein production in modeled information with different levels of P5 inhibition. The table displays the ratio of phage species accumulated in the cell at the end of 60 minutes post infection when P5 inhibition on translation of proteins P1, P2, P3, P5, and P10 turned off and divided by the amount of accumulated species at the end of 60 minutes when it is turned on. * Indicates similar ratios measured in-vivo by the following sources [30, 31]

Case	Phage	dsDNA	P5DNA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
All Proteins	0.00	0.02	0.00	2.80	21.30	3.52	0.90	2.11	5.11	2.29	10.95	0.85	123.04	1.77
Only on P2 Translation	0.77	1.20	1.80	0.78	55.14	1.10	0.77	1.72	0.55	1.84	1.63	0.80	1.45	0.73
Only on P10 Translation	0.00	0.02	0.00	1.24	0.42	0.70	1.24	0.17	10.36	0.15	15.47	1.54	219.13	2.52
Only on P5 Translation	0.51	0.19	1.85	0.41	0.95	0.13	0.78	4.26	1.78	4.75	0.03	0.52	0.89	1.21
Only on P3 Translation	1.00	1.01	1.01	0.98	1.00	2.38	0.97	1.01	0.96	1.01	0.47	0.97	1.00	0.97
Only on P1 Translation	1.00	1.00	1.00	1.21	1.00	1.00	1.00	1.00	1.00	0.99	0.97	1.00	1.00	1.00
Experiental Results*	N/A	N/A	N/A	6.5	20.2	18.7	N/A	3.6	N/A	N/A	N/A	N/A	9	N/A

Table 15 shows the results of the model experiments described in the preceding paragraphs and the experimental results obtained from literature [30, 31]. Completely removing all inhibition produces no phage within the model because DNA synthesis does not occur because P10 is made at a greater level than P2. Having more P10 molecules than P2 inactivates P2 function preventing the conversion of RF1 to RF2 DNA, an essential step in DNA replication. No phage production or DNA production occurs when inhibition of P10 translation is turned off. According to the model, control over P10 translation appears to be an essential aspect of phage replication. The model includes an unrealistically high binding constant for P10 binding to P2 to assure that excess P10 shuts down phage production. Additionally, the model was parameterized such that P5 has a greater effect on P10 translation than is indicated by the experiments. The ratio of P10 production without and with P5 is predicted to be 123.04 (when P5 translation is turned off for all proteins) or 219.13 (when P5 translation is turn off for only P10) which is over 10 times greater than the experimentally determined P10 inhibition. A likely answer to the

discrepancy is an over-estimate of the ribosome binding site strength of P10 because a secondary structure exists by the ribosome binding site which blocks the ribosome [19].

In the next couple of paragraphs “wild-type infection” will indicate the model results where all P5 inhibition is active. Removing P5 inhibition on the translation of P2 reduces phage production by 23 percent of wild-type levels. The decrease occurs because less assembly sites are being made which is a direct result of a decrease translation of P1, the limiting factor in assembly sites. The number of P1 molecules produced at the end of 60 minutes post infection decreases 22 percent. The decrease in P1 production occurs because more ribosomes are devoted to translating P2 when P5 does not inhibit P2 translation, see Figure 39. P2 encoding mRNAs are more prevalent and have a stronger ribosome binding site and are thus preferentially translated over P1 genes.

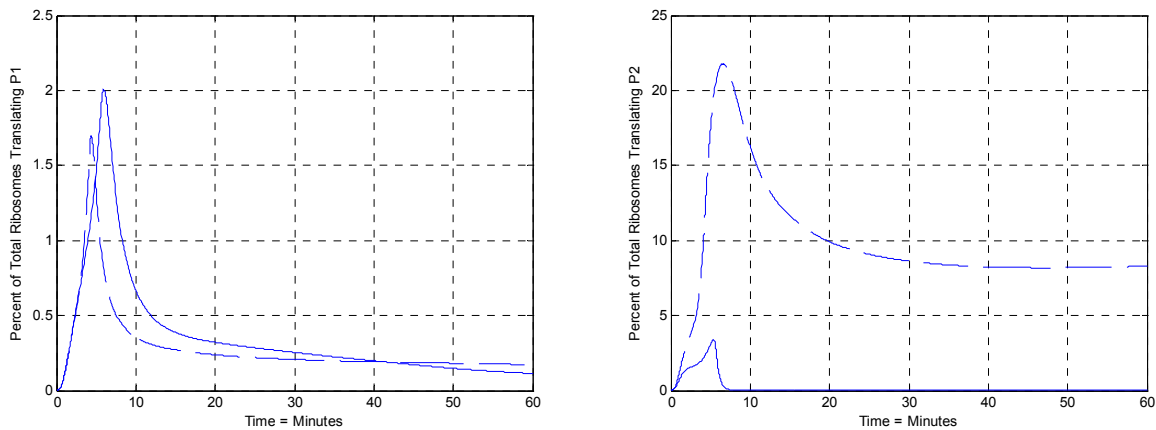


Figure 39 – (Left) compares the percent of ribosomes actively translating P1 as a function of time when inhibition on translation of P2 is turned on (solid line) and turned off (dashed line). (Right) compares the percent of ribosomes actively translating P2 as a function of time when inhibition on translation of P2 is turned on (solid line) and turned off (dashed line).

Removing the P5 inhibition on its own translation decreased the phage production at the end of 60 minutes by a factor of 2 compared to wild-type infection. The decrease in phage production was directly caused by the decreased amount of P8 molecules being created during

the infection process. When P5 translation inhibition is turned off, approximately 50 percent of all ribosomes are dedicated to translating P5 while only 22 percent are involved in the translation of P8, as seen in Figure 40. Translation of P5 is dominant because it has a stronger ribosome binding site than P8. When P5 inhibition is turned on, less than 5 percent of ribosomes are translating P5 and 55 percent are translating P8. According to the model P5 must inhibit its own translation to allow enough of the allocated cellular resources to make P8 molecules throughout the infection, otherwise P8 becomes the limiting component in phage production.

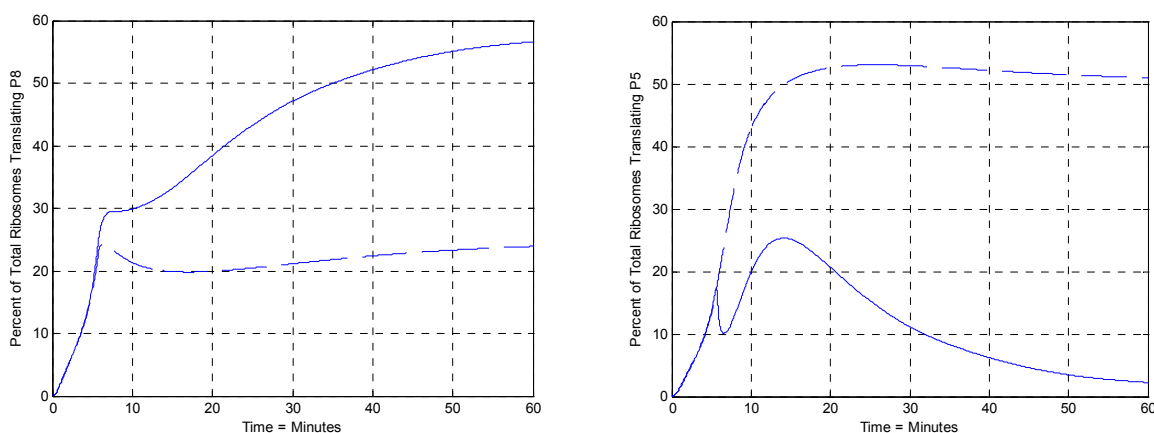


Figure 40 - (Left) compares the percent of ribosomes actively translating P8 as a function of time when inhibition on translation of P5 is turned on (solid line) and turned off (dashed line). (Right) compares the percent of ribosomes actively translating P5 as a function of time when inhibition on translation of P5 is turned on (solid line) and turned off (dashed line).

The relative binding site strength of P5, as predicted by the ribosome binding site calculator created by Salis lab, is the strongest and was used as the basis for calculating relative promoter strengths. In the model, the P5 ribosome binding site was assigned a promoter strength of 1.00. While the relative binding site strength of P8 is weaker at 0.22. At first glance the numbers appear to be incorrect because P8 is needed in the highest quantity followed by P5. However, the relative ribosome binding site strengths may be reflective of the true value and may be part of the phage design. Early in the infection P5 must be made quickly in order for

dsDNA production to decrease and P5DNA production to increase. P5 is required in high amounts because it requires 1600 molecules of P5 to sequester a single ssDNA genome. Additionally, the assembly sites are still being formed early in the infection and phage assembly is not running at full capacity. When phage assembly is not running at full capacity P8 translation is not needed at a high rate. Late in the infection as the number of assembly sites increases, the rate of P8 packaged onto nascent phage increases, and to meet the demand, P8 translation must increase. To meet increased P8 demand, P5 translation is turned off and additional ribosomes can be recruited to translate P8. P5 translation does not need to occur at high levels late in the infection because it is continuously recycled during the assembly process from P5DNA. In other words, P5 is needed first in the infection in order for P5DNA formation to begin and to control the total copy number of replicative form DNA in the cell, and P8 is needed late in infection for phage assembly.

Turning off the P5 inhibition on the translation of either P1 or P3 is predicted to have the least effect on the overall phage infection, see Table 15. The increase in the number of molecules of either protein is not significant in the absence of P5 translational repression. P1 shows an approximate 3 fold increase and P3 shows an approximate 4 fold increase in concentration when P5 inhibition is removed. Both proteins have minimal effects on the outcome of phage synthesis because neither uses a significant portion of the ribosomes available for phage replication (see Figure 41). Translation of P1 is probably controlled by P5 because overproduction of the protein can kill the host cell by inserting too many pores within the cell membrane [5]. The reason why P3 is inhibited by P5 may be to divert more ribosomes to the translation of P8 because P3 does have a stronger ribosome binding site than P8. Table 15 shows

a decrease in the amount of P8 remaining in the cell compared to wild-type phage infection when P3 translation inhibition is turned off.

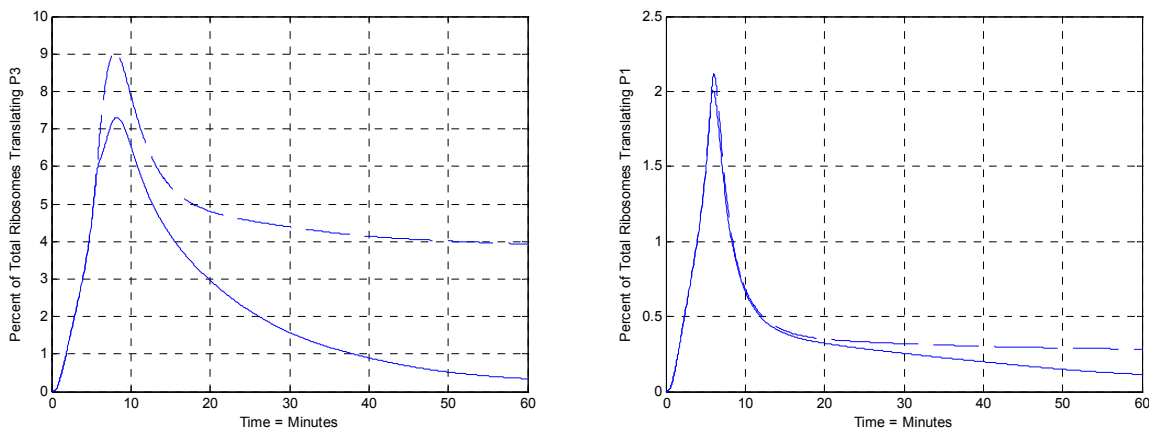


Figure 41 - (Left) compares the percent of ribosomes actively translating P3 as a function of time when inhibition on translation of P3 is turned on (solid line) and turned off (dashed line). (Right) compares the percent of ribosomes actively translating P1 as a function of time when inhibition on translation of P1 is turned on (solid line) and turned off (dashed line).

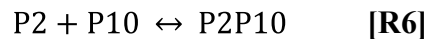
A further study from Konings laboratory has stated the inhibition of P2 by P5 is a dispensable function and no observable difference in phage production is measured [87]. When removing P5 inhibition in the model, it predicts phage production to decrease by 23 percent. The 23 percent difference is probably within error of experiment and can be considered not have an effect on phage production. The caveat of the experiment is in order to eliminate P5 inhibition they deleted nucleotides at the beginning of the sequence to prevent P5 inhibition. The act of removing these nucleotides may have decreased the ribosome binding site strength and downplayed the importance of P5 inhibition. It would have been beneficial if Zaman et al. [87] provided a comparison of P2 production between the two infections because over production of P2 is suggested to have an negative effect on P5DNA formation and the mechanism is unclear. The proposed mechanism is discussed in the next section.

3.4 Additional Control Over dsDNA Synthesis

3.4.1 Background Information

The way the current model is set up, P5 is the only protein which can decrease the rate of dsDNA synthesis and increase the rate of P5DNA formation. By varying the rate of P5DNA formation (by changing the rate constant C_7 in equation E7) we were able to tune the accumulation of dsDNA. The way the current model is simulating the switch from dsDNA synthesis to P5DNA synthesis is very simplistic to what is actually going on in the cell because it is only a function of P5 accumulation which is contradictory to conclusions made by Fulford and Model [14, 15]. They concluded the accumulation of dsDNA and P5DNA is dependent on three phage proteins P2, P5, and P10.

Fulford and Model were able to test four separate scenarios which had only two phenotypic outcomes [14, 15] . In the first scenario P2 translation was inhibited and DNA replication did not occur. The first scenario was expected because it was already well-known viral protein P2 was required to nick the dsDNA to initiate replication. In the second scenario the overproduction of P10 caused the DNA replication to stop. The pronounced effect of P10 on replication was a surprise for Fulford and Model. Scenarios 1 and 2 produce the same result for different reasons. Fulford and Model postulated that P10 inhibits P2 by direct binding. The complex formed when P2 and P10 bind inhibits P2 function. We postulated P2 can bind to P10 in a one to one ratio producing an completely inactive complex P2P10 complex. We implemented the mechanism in the model and are able to observe no DNA replication when P10 was overproduced.



In the third scenario tested by Fulford and Model, P2 was overproduced relative to wild type levels. The result of overproducing P2 reduced the accumulation of P5DNA and increased the accumulation of dsDNA (all forms of dsDNA), even in the presence of P5. Indicating P2 may have an additional role of having a positive effect on the complimentary strand synthesis. Two possible mechanisms were suggested. The first mechanism had P2 promoting complimentary strand synthesis directly by recruiting RNA polymerase more efficiently than P5 sequestering DNA. The second mechanism had P2 protecting the (-) strand origin of replication, preventing P5 from binding to that region of ssDNA. In the last scenario, no P10 was produced and the same result was observed as if P2 was overproduced.

To summarize the results discussed in the last two paragraphs, P2 and P10 appear to have roles in the synthesis of ssDNA, and in the accumulation dsDNA and P5DNA. The role of P2 and P10 on the accumulation of ssDNA formation is relatively simple and can be graphically represented in Figure 42A, where P2 has a positive effect on the formation of ssDNA and P10 has a negative role. The factors effecting the accumulation of dsDNA and P5DNA synthesis is complex and the graphical description in Figure 42B tries to aid in the description of the effects. Looking at Figure 42B it appears biology may set up a complex mechanism to ensure dsDNA accumulate in appropriate amounts to ensure maximum efficiency of phage replication.

The exact mechanism for how P2 and P10 can affect the accumulation of dsDNA and P5DNA accumulation has not been studied. I believe two equally possible mechanisms exist. The first mechanism, P10 is required to melt the (-) strand origin of replication because P5 cannot do it by itself. In the second mechanism, P2 binds to the (-) strand origin of replication and protects it from being fully sequestered by P5 molecules. In the next couple paragraphs the two mechanisms are expanded.

of replication which consists of hairpins B and C where any perturbation to the (-) strand origin region reduces DNA synthesis [4].

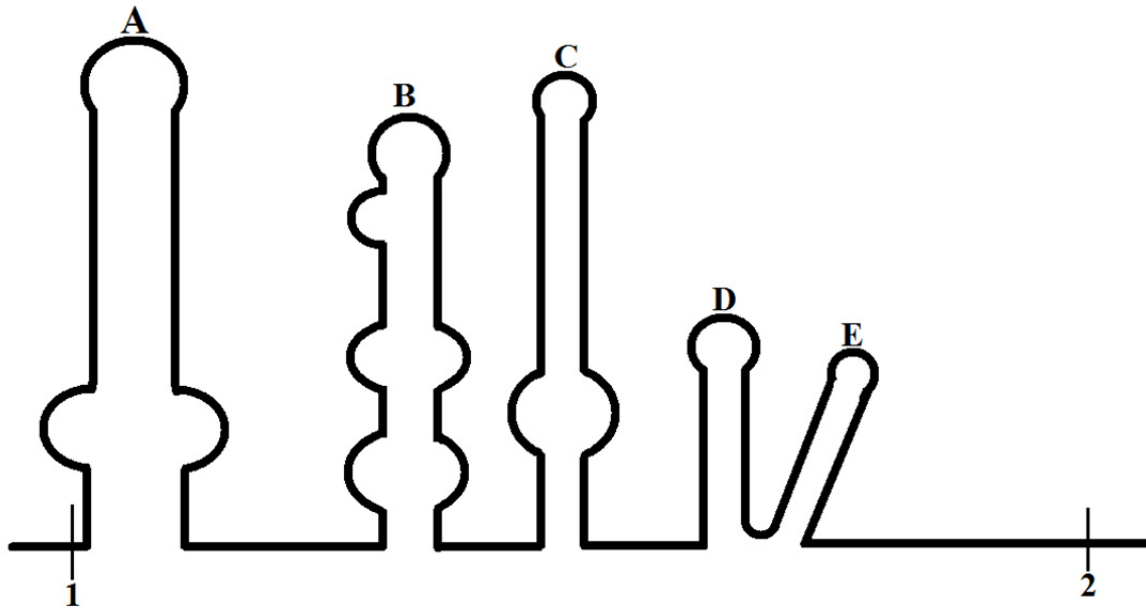


Figure 43 – A schematic drawing of the secondary structure of the intergenic region of the ssDNA. Each hairpin is marked with a letter A-E where each has a specific role in the infection process. Hairpin A is the packaging signal and is needed for efficient assembly of phage particles. Hairpins B and C are required for the conversion of ssDNA to dsDNA. The regions containing the hairpins D and E are required for DNA replication to occur. The location indicated by 1 indicates the end of gene 4 coding sequence and the location indicated by location 2 is the indication of the beginning of gene 2 coding sequence. The schematic is an adaption from the following source [88].

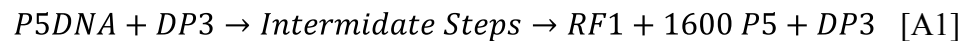
If we assume P5 cannot disrupt hairpins B and C to sufficiently prevent (-) strand synthesis, it would support the hypothesis P10 is required to either prevent dsDNA synthesis from occurring or disrupting hairpins B and C to allow full sequestration of the ssDNA. In order for this mechanism to fit all four scenarios Fulford and Model tested, both P2 and P10 must exist in the cell as free molecules, which the model currently does not have any free P10 accumulated in the cell. However the presence of free P2 or P10 in the cell must be very sensitive of the relative ratios which exist between the two. If P10 is overproduced, it must be

able to shift the equilibrium to the reactant side where P10 is bound to most of the P2 in the P2P10 complex. If P2 is overproduced it must be able to bind to most of the P10 molecules to inactivate the function of P10 on the ssDNA intergenic region.

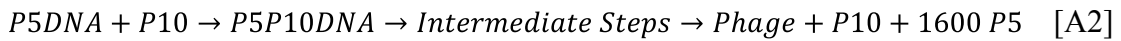
In the second mechanism, we assume P5 has the ability to melt hairpins B-E and does not require another protein for full sequestration of ssDNA. Suggesting P2 may have a direct role in protecting hairpins B and C. When P2 is produced in wild type levels, it cannot protect hairpins B and C before P5 completely takes over the ssDNA. If it is overproduced, the binding reaction can occur faster than complete sequestration of P5. P10 is thought to have a function of damping the effects of P2 by disabling active P2 from the cell. By this mechanism, the wild-type levels of P2 are enough for dsDNA synthesis to be dominant. However, P10 is produced to dampen the effects of P2 function. Since both of the discussed mechanisms are pure speculation we believe both have an equal probability of being proven or disproven.

3.4.2 P10 Binding to ssDNA Mechanism

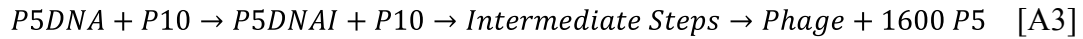
In the proposed mechanism, the model will still assume 1600 molecules of P5 can bind to ssDNA in one step, however the (-) strand origin of replication will remain intact. Meaning P5DNA may be converted to dsDNA. An additional set of reactions, A1, must be added to the model to account for the conversion of P5DNA to RF1 DNA. It is assumed the elongation of DP3 will not be affected by P5 molecules bound to the system. The P5DNA is assumed to still be able to bind to the cellular membrane to initiate phage assembly; however it will be severely hindered because ssDNA conversion to dsDNA will occur much faster than phage assembly initiation.



In order for a P5DNA to be fully protected from conversion to dsDNA, I am predicting P10 must interact with hairpins B and C to disrupt the origin of replication. There are at least two possible mechanisms through which P10 could inhibit dsDNA production. The first mechanism involves P10 actively binding to the origin of replication and preventing RNAP initiating (-) strand synthesis. A2 is the reaction describing the mechanism where P5P10DNA is P5DNA bound to one P10 molecule which inhibits dsDNA synthesis.



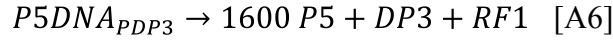
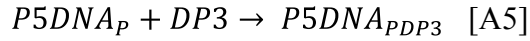
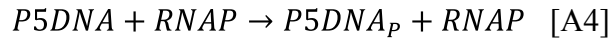
Another option is that P10 acts as a catalyst to disrupt hairpins B and C and to allow P5 to bind to those regions. The proposed mechanism is as follows where P5DNAI is ssDNA bound to 1600 P5 molecules, and hairpins B and C are not intact:



The current proposed mechanism for P10 function relies heavily on another proposed reaction R6. Future experiments should be aimed to determine if P2 and P10 have any binding affinity to each other and if they do determine kinetic parameters. If free P10 is required for wild-type phage infection, then the current model is over-estimating the binding affinity P2 and P10 because no free P10 exists in the model even when wild-type levels of P2 are made.

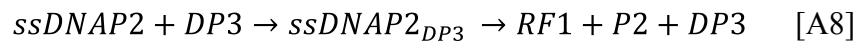
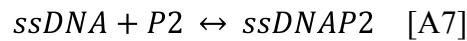
The DNA replication must also be modified slightly to account for the new modifications. The proposed function of P10 relies heavily on (-) strand replication occurring faster than phage assembly initiation. In the current model, it would be a competition of P5DNA binding to DNA polymerase 3 (DP3) or an assembly site. However, DP3 is only produced to 10-20 copies per cell [60]. It may be more accurate to model the rate of RNA primer formation by RNAP. RNAP binding will occur more frequently because about 1000 times more RNAP is

present in the cell than DP3 and RNAP will have a significant impact on the conversion of P5DNA to dsDNA. The new model would account for the RNA primer formation of the P5DNA molecules but will not account for elongation of the primer and termination. It will also assume that when the primer is created it cannot be packaged into phage particles. Equations A4-A6 describe this model where $P5DNA_p$ indicates P5DNA with a primer bound to it and $P5DNA_{pDP3}$ is an P5DNA bound to a DP3 enzyme.



3.4.3 P2 Protecting the (-) Strand Origin of Replication Mechanism

To implement the P2 protecting the (-) strand origin mechanism in the model only requires two additional equations which accounts for the forward and reverse reaction of P2 binding to ssDNA. The binding reactions are described in equation A7 and A8 where $ssDNAP2$ is ssDNA bound to 1 P2 molecule and $ssDNAP2_{DP3}$ is ssDNA bound to both P2 and DP3 molecule.



In order for the mechanism to work, reaction A7 equilibrium must favor the unbound P2 (reactant side) of the equation when wild-type levels of P2 are made. If P2 is overproduced then reaction A7 equilibrium would shift towards the right, to bound ssDNA.

3.4.4 Future Experiments

The two mechanism discussed in the section are purely speculative and have no direct experimental evidence. Therefore, we must experimentally test in the lab if either mechanism

exists. We could test the mechanisms in-vitro by first isolating viral ssDNA, P2, P5, and P10 separately. Then we would conduct three separate tests.

In the first test, we will mix ssDNA and all components supplied by the cell for DNA replication (RNA polymerase, DNA polymerase III, DNA gyrase, nucleotides, and etc.) together in one vial. The first test will serve as the control because we should see all ssDNA being converted to dsDNA. Rolling circle replication will not occur because P2 is not present to nick the DNA.

In the second test, we will first mix ssDNA and enough P5 to sequester all ssDNA in a tube and allow it to equilibrate. Then we will add the components required for DNA replication. If we observe dsDNA formation then it would suggest P5 cannot disrupt the origin of replication by itself and requires an additional protein. If no dsDNA synthesis occurs, it suggests P5 is sufficient to sequester ssDNA and provides evidence against the proposed mechanism involving P10 to disrupt the origin of replication. One potential pit-fall to the current test is we may produce an environment which has an unrealistically high ratio of P5 to ssDNA and P5 is overcoming the hairpins because free P5 is more concentrated in-vitro than in wild-type infection in-vivo.

If the second test produces dsDNA then the test suggests that P5 is not sufficient to prevent dsDNA synthesis, a third test can be performed to see if P10 has a direct role in the sequestration process. In the third test, we would mix ssDNA, P5, and P10 in a vial and allow it to equilibrate. Then we would add DNA replication components. If the ssDNA is prevented from converting to dsDNA then it would indicate P10 has a direct role in P5DNA formation.

Otherwise, it would indicate an additional protein besides P10 is needed to prevent complimentary strand synthesis.

We will test if P2 can promote complimentary strand synthesis directly by protecting the (-) strand origin of replication. We will first mix ssDNA and P2 in a tube and allow it to equilibrate. Then we will add enough P5 to sequester all ssDNA and allow the tube to equilibrate again. Finally, we will add the replication components. If P2 has a direct role in preventing P5 from melting the (-) strand origin of replication then most of the ssDNA should be converted to dsDNA.

3.5 Cell Replication

All of the previous discussions of analyzing the model simulations have only analyzed the infection up until the first cell division. We can extend the simulation to account for cell replication by running the simulation again but changing the initial number of species at the start of the new simulation. We assumed when cell division occurred at the end of the cell cycle, all model species were evenly distributed between the two daughter cells. For example, if 300 mRNA strands were made at the end of the cell cycle, each daughter strand will contain 150 each. Therefore the initial amount of species at the beginning of the new cell cycle will be set to half of the value of the final amount of species made at the end of the previous cell cycle.

When all cell species are equally divided it also include the host resources RNA polymerase (RNAP), DNA polymerase 3 (DP3), and ribosomes. Since the model does not simulate the synthesis of these enzymes produced by the cell, we must add enough enzymes at the start of the new simulation (or cell cycle) so the total number remains constant throughout each cell cycle. It turns out we must add $\frac{1}{2}$ of the initial value used at the start of the first cell cycle. Figure 44 is a visual description of the simulation strategy for cell division.

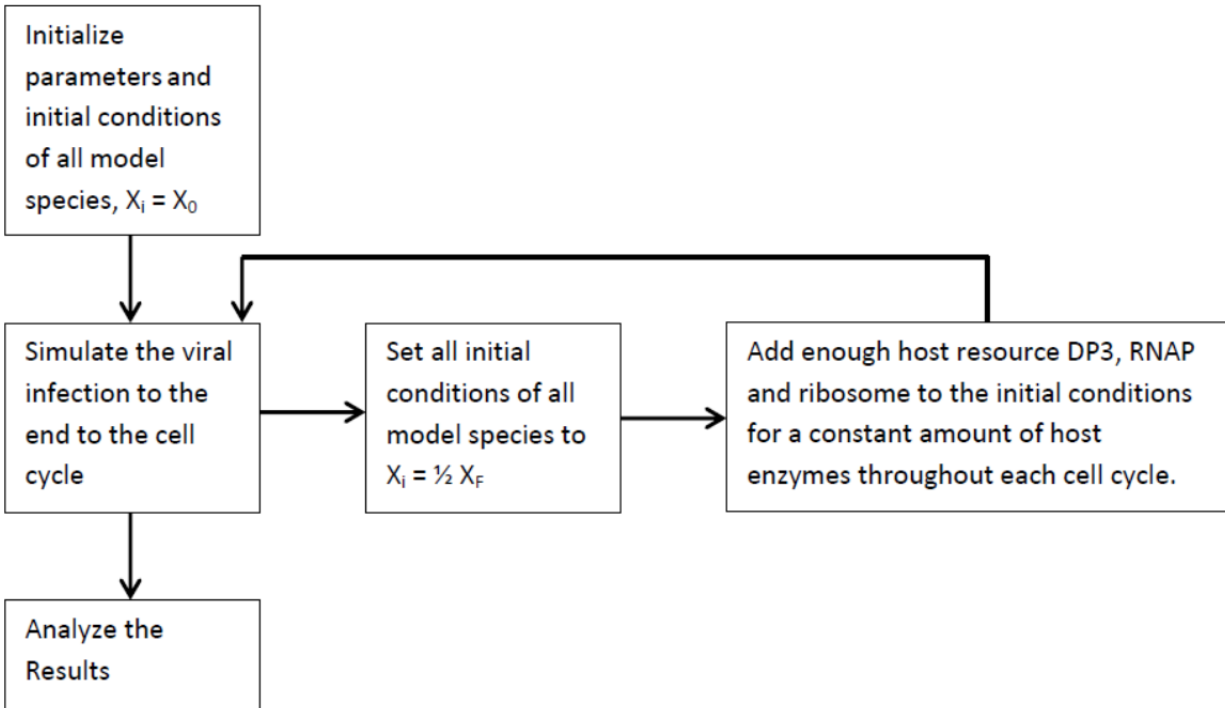


Figure 44 – The simulation strategy for simulating phage infection through cell division. X_i indicates the initial number of model species at the start of the cell cycle. X_0 indicates the initial number of model species at the start of the first cell cycle. X_F indicates the final number of phage species at the end of the cell cycle.

When 10 cell divisions are simulated using the model strategy as described above, the rate of phage produced from a single cell decreases after 5-6 cell replications, see Figure 45. *E. coli* cells infected with phage have been shown to have the ability to “cure” themselves after 7 - 15 generations, meaning stop producing phage [10, 89, 90]. Therefore the model and experimental “curing” compare favorably. It should be noted that the model was not directed in any kind to produce the following result. We can consider this a true prediction from the model because the model was built by analyzing the first hour after cell division. When Figure 45 was first generated we initially thought the model might be wrong because we were initially expecting a steady-state to occur.

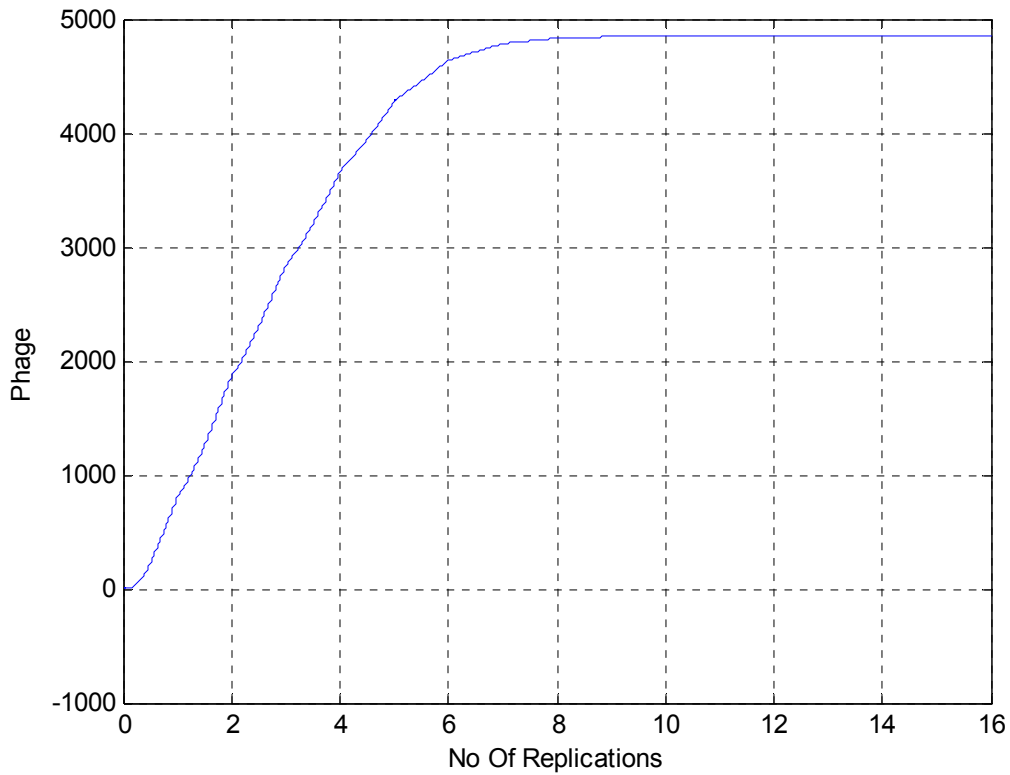


Figure 45 – The total number of phage produced by a single cell through multiple generations of cell division. Each cell division was assumed to take 60 minutes to complete.

According to the model, phage replication stops after the 6th cell division because the total number of dsDNA ($RF1 + RF2 + RF2/DP3$) accumulation decreases with every cell division, see Figure 46. In the first hour after infection (before the 1st division) 50 strands of dsDNA are synthesized in the cell. For the next four cell cycles, dsDNA accumulates in the cell at approximately 10 strands per hour. Then after the fifth division it decreases to approximately 3 strands per hour and eventually to 0 strands per hour. The observed decrease in the rate of dsDNA synthesis directly correlates with experimental observations and conclusions [10].

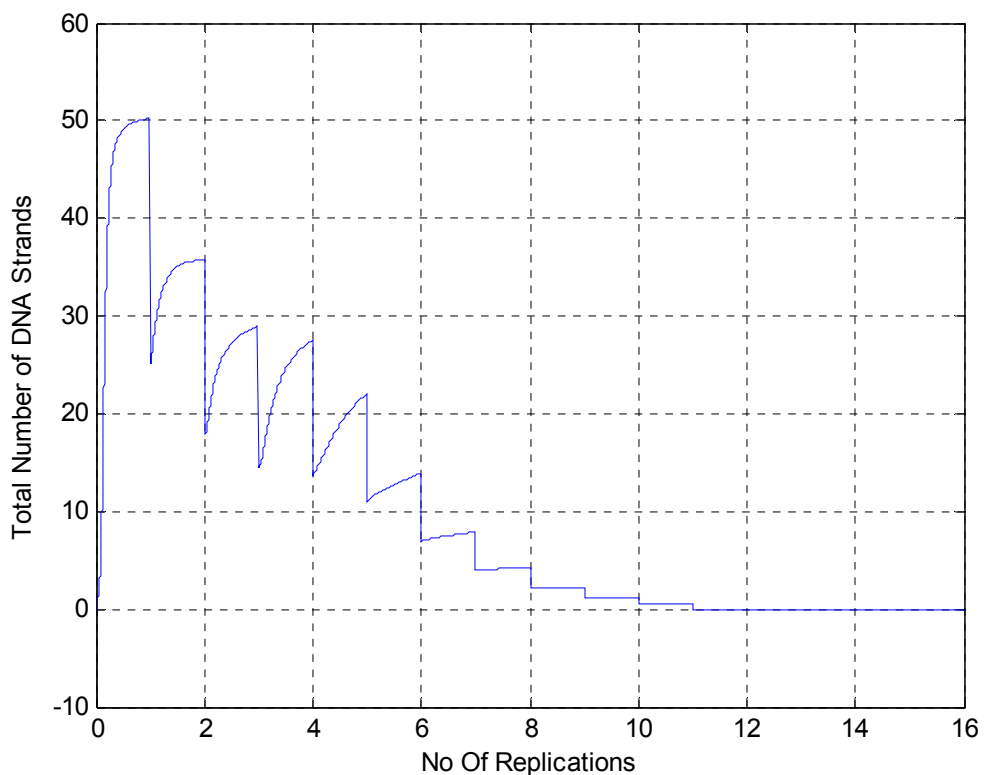


Figure 46 - The total number of dsDNA (RF1 + RF2 + RF2/DP3) accumulated in the cell through multiple generations of cell division. Each cell division was assumed to take 60 minutes to complete.

According to the model, accumulation of dsDNA is inhibited because P5 accumulation in the cell is maintained at high concentrations (greater than 10^4 molecules) throughout 15 cell divisions, see Figure 47. Since P5 is made at a high concentration most of the ssDNA will be converted to P5DNA. Our conclusions directly parallel to experimental conclusions made by another group [89]. The total amount of P5 increases after the 10th cell division increases and it is unclear on why this occurs.

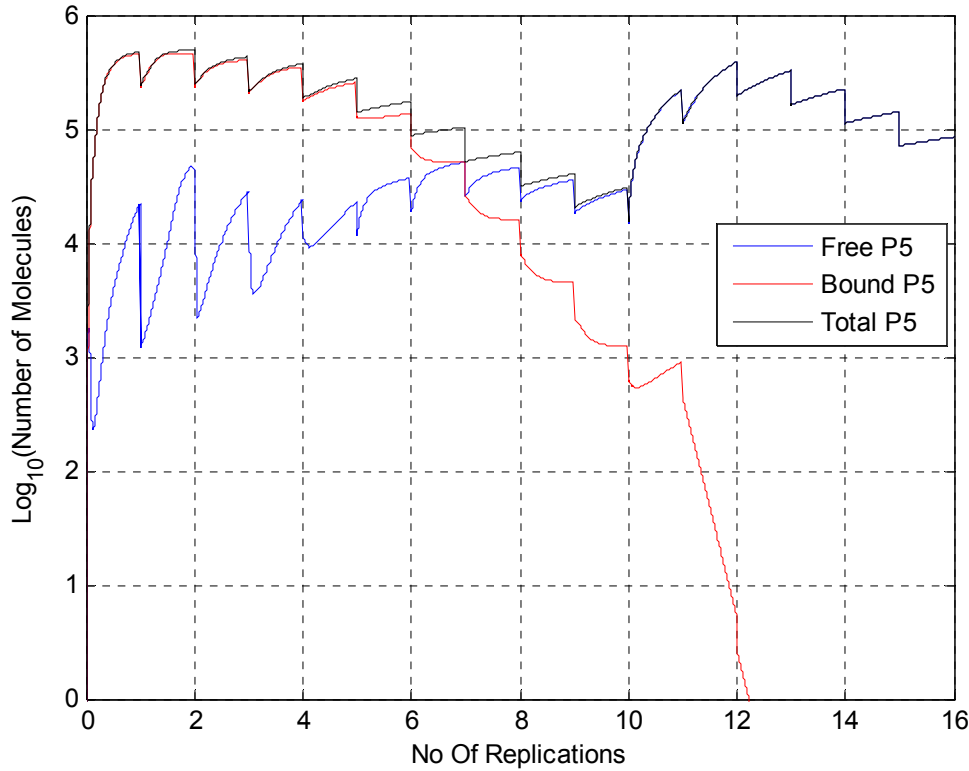


Figure 47 – The accumulation of all forms of P5 throughout 15 cell divisions. the y-axis is a log based 10 graphs of all P5 species within the cell. The blue line indicates free P5 molecules, the red line indicates all P5 that is currently bound to ssDNA, and the black line indicates total P5 (bound and unbound P5 molecules). Each cell division was assumed to take 60 minutes to complete.

3.6 Overestimation of the Rate of Phage Production

A key model parameter, rate of phage assembly (C_{19}), was estimated through the accepted value of 1000 phage produced in the first hour post infection. Additionally, all of the P5 efficiency factors were also indirectly based off of C_{19} . Many articles and review articles mentioned the number but do not site the source. A recent study has observed a burst size of 250 phage particles [81] and we believe this number to be more accurate than 1000 phage burst size. Therefore, we are going to briefly discussed how the model changes when we calculate C_{19} according to the 250 phage produced in the first hour. The new calculation is as follows:

$$C_{19} = 250 \frac{\text{Phage}}{\text{Hour} * \text{Cell}} * \frac{1 \text{ Hour}}{3600 \text{ Second}} * \frac{1 \text{ Cell}}{225 A_s} * \frac{1 A_s}{1 P_l} = 3.1 * 10^{-4} \frac{\text{Phage}}{\text{Second} * P_l}$$

When we reduce the rate of elongation in the model multiple changes occur to the accumulation of model species at the end of 60 minutes post infection and are summarized in Table 16. When the rate of phage synthesis decreases the total amount of phage produced at the end of 60 minutes also decreases. Since less phage are being produced, the accumulation of all phage proteins P3, P6, P7, and P8 (exception of P9) in the cell increase as well. The coat proteins accumulate in higher quantities because less is being exported by the cell into newly formed phage. The reason why P9 accumulation decreases is not clear from Table 16. The accumulation of P5DNA increases by a factor of 2.5 because it is not being exported as quickly. Since more P5DNA accumulates in the cell, less unbound P5 molecules accumulates. With less unbound P5 molecules in the cell, the effect of P5 on translation of its own molecules decreases and thus why the total number of P5 molecules (bound + unbound) in the cell increases.

With less free P5 accumulated in the cell we would expect an increase in the number of P1, P2, P3, and P10 also. We do see a small increase in P2 and P10, and a large increase in P3. The increase in P3 accumulation occurs for two reasons; first P5 inhibition on translation decreases because of the decrease in the amount of P5 and it is not being exported in newly assembled phage as quickly. We do see a small decrease in the amount of P1 produced and it is uncertain on the exact reasoning behind it. It is most likely due to how the ribosomes are distributed for translation.

The rate of phage elongation does have a significant effect on the outcome of the model. When we decrease the rate of phage assembly the model does not do anything unexpected (we are considering the decrease in P1 accumulation to be minor). Therefore the model may be over-

estimating the amount of phage produced in one hour but we believe the rate of protein synthesis correlates well with biology, see section 3.1 for more details.

Table 16 – Compares the final number of model species at the end of a 60 minute cell cycle when changing the rate of phage elongation, which is parameter C_{19} in the model. All numbers are in units of molecules.

Species	1000 Phage Produced in 1 hour	250 Phage Produced in 1 Hour
Phage	8.05E+02	1.93E+02
dsDNA	5.02E+01	9.15E+01
P5DNA	2.98E+02	8.34E+02
P1	4.17E+03	3.80E+03
P2	1.75E+03	1.80E+03
P3	1.59E+04	2.43E+04
P4	2.29E+05	1.71E+05
Unbound P5	2.21E+04	6.44E+02
Total P5	4.76E+05	1.34E+06
P6	3.89E+03	5.42E+03
P7	4.21E+04	1.31E+05
P8	1.17E+05	1.21E+06
P9	9.11E+05	6.92E+05
P10	4.79E+02	5.38E+02
P11	1.01E+04	7.61E+02

3.7 Expanded Pathway for mRNA Degradation

As suggested in the section “Validation of the Model”, the mRNA degradation pathway for the more frequently transcribed region may not be sufficiently representing the biology. The discrepancies include inconsistent ratios of mRNA levels in comparison to mRNA H and the time at which steady-state is reached for a number of the phage components is also inconsistent with experimental results.

The model is currently using a very simple cascade mechanism where each mRNA species is consecutively degraded into smaller species, as shown in the red arrows in Figure 48. However the actual mechanism is most likely more complicated than what is implemented in the

model. First, the model completely ignores the degradation pathway which forms mRNA C. Secondly, mRNA transcript degradation have been studied in-vivo and half-life experimentally determined, but we do not know the actual pathway. For example, in the model mRNA A will be quickly degraded into mRNA D and subsequently mRNA E, F, G, and H by endonuclease activity. Additionally, at each step for mRNA D, E, F, and G, they may be degraded in one step by an exonucleases activity. The exonucleases activity is thought to only account for 30 percent of degradation [77] and is considered the minor pathway of degradation. For example, most mRNA A will degrade through six steps before the complete degradation of the mRNA is accomplished. Shorter paths may exist because endonucleases may bind multiple sites on the mRNA, for example $A \rightarrow E \rightarrow H$ as shown in Figure 48. Bypassing the mRNA D and F will reduce the ratios of D to H and F to H and it may make Table 14 more consistent. The only problem with the expanded model is more information would need to be known to know which pathways are used more frequently than others.

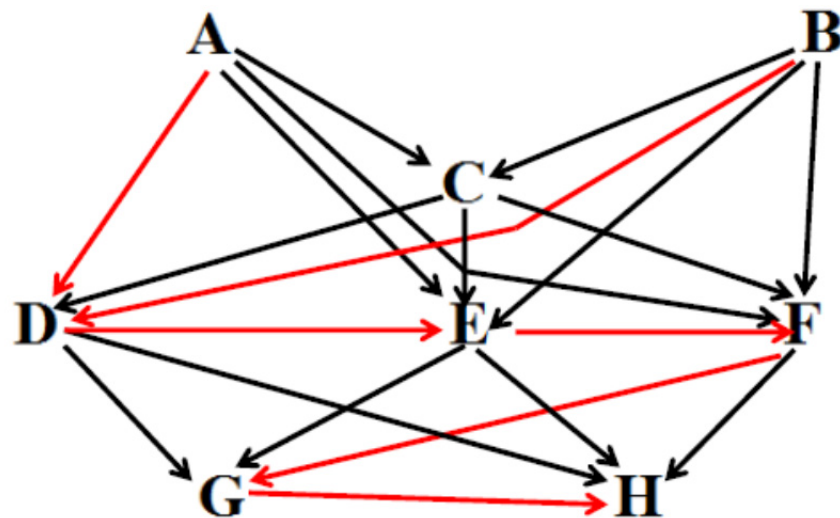


Figure 48 – A visual representation of all of the pathways an mRNA from the more frequently transcribed region may undertake. The red arrows represent the pathway the current model is using. It is also assumed each mRNA can be completely degraded in one step by an exonucleases enzyme but it is not explicitly shown in the figure.

3.8 Possible RNAP Interference

Each promoter site within the model acts independently from each other and the transcription of mRNA A cannot interfere with the transcription of mRNA B. Due to the overlapping structure of the phage genome, interference is highly probable and may have significant impacts on the outcome of mRNA synthesis. However, adding interference would require keeping track of the location of each DP3, and the generation and consumption of the interfered promoter sites. The complexity of the model and the computational cost will significantly increase and was therefore ignored in the current model. A future model could account for interference to study the possible effects it has on viral infection.

Chapter 4: Conclusion

The current model has been created from information already published in literature. The model was based as much as possible on experimentally determined constants found from literature, however some experimentally undetermined kinetic parameters were fitted to achieve experimental observation. The only validation of the model exists from comparing results from a wide range of resources. The comparisons were not limited to comparing simple aspects such as comparing the amount of dsDNA accumulated in the cell as predicted by the model and published experimental observations. Further validation of the model should continue and the unknown kinetic parameters should be experimentally determined.

The model potentially gave us insights into the process as a whole and may have shed light on some unclear mechanisms. First it potentially gave us insight into how the host resources are utilized. Secondly it suggests possible meaning to why P5 inhibits translation of P1, P2, P3, P5, and P10. For example, the model predicts early in the infection process 25 percent of allocated ribosomes are dedicated to making P5 and late in the infection the percentage drops to 5, see Figure 36. The decrease in the rate of translation late in the infection is a consequence of P5 inhibiting its own translation. The inhibition of its own translation may be important in order for P8 to be made in sufficient quantities for phage elongation. The reason why the phage may have evolved to inhibit P5 translation rather than decrease the ribosome binding site strength is to allow P5 to be made in high quantities early in the infection and translated slow late in the infection. Early in the infection a large pool of P5 molecules is needed to convert dsDNA synthesis to P5DNA synthesis and once the pool of P5 molecules is made, it is continuously recycled from P5DNA being extruded from the cell. Lastly, in the process of building the model, the P2 and P10 binding interaction was implemented, which has not been directly observed

experimentally, and a proposed mechanism P10 inhibiting P5DNA formation even in the presence of P5 in the cell is proposed.

There are still unknowns about the biology which include the exact mRNA degradation for the frequently transcribed regions. Also the P2 and P10 interaction is purely speculative and needs to be verified in the lab. During the writing of the paper, new information continues to be found which would modify the model. For example, the number of phage produced in one hour post infection was assumed to be 1000 phage particles, which is an accepted value within the community and mentioned in at least two review articles [3, 91]. However two papers suggests the number may be closer to 200 phage particles produced in the first hour [10, 81]. We are suspicious of the 1000 phage produced in one hour because both review articles do not site a source. The outcome of the new information would change the rate of P8 addition in the assembly process and increase the number of P5DNA molecules in the cell not being actively assembled.

Future iterations of the model should connect the phage infection and *E. coli* cell biology. Currently the only interaction between the cell and phage infection taken account in the model is limiting cellular resources. The current model implicitly assumes phage infection the host cells volume remains constant. *E. coli* cells are continuously growing and the volume is changing, therefore the change in volume may affect the rate of some reactions in the model. Secondly, the host cell resources are assumed to be a static number. However in a growing cell, the cellular resources are going be changing as the cell is growing, presumably increasing resources such as RNA polymerase, DNA polymerase III, and ribosomes. Future models should analyze how variable host resources throughout the infection affect phage infection.

The model described in this paper is a deterministic simulation, which means for a given set of inputs the exact same result will occur every time. However, virus infection is very noisy and does not produce the same number of molecules at the end of 60 minutes post infection. For example, if we assume 1000 phage can be produced in the first hour post infection then some cells may produce 600 particles and some 1200 particles. The variation in the amount of phage production is most likely due to the intrinsic fluctuations of phage species during the infection process. We can study which species have the most effect on the final phage production by using a stochastic approach in simulating the model, which has been done before for the vesicular stomatitis virus [55]. A stochastic approach introduces a random variable into the simulation and for any given set of inputs a different result will occur. The biggest drawback on using a stochastic approach is that it takes a significant amount of computational power to simulate.

References

1. Ackermann, H.-W., Classification of Bacteriophages in The Bacteriophages R. Calendar, Editor 2006, Oxford University Press: Oxford ;.
2. Webster, R.E., Filamentous Phage Biology, in Phage display a laboratory manual, C.F. Barbas, Editor 2001, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
3. Russel, M. and P. Model, Filamentous Phage in The Bacteriophages R. Calendar, Editor 2006, Oxford University Press: Oxford ;.
4. The Bacteriophages. Viruses, ed. R. Calendar 1988, New York: Plenum Press.
5. Horabin, J.I. and R.E. Webster, MORPHOGENESIS OF F1 FILAMENTOUS BACTERIOPHAGE - INCREASED EXPRESSION OF GENE-I INHIBITS BACTERIAL-GROWTH. *Journal of Molecular Biology*, 1986. **188**(3): p. 403-413.
6. Lafarina, M., TRANSCRIPTION IN BACTERIOPHAGE-F1-INFECTED ESCHERICHIA-COLI - VERY LARGE RNA SPECIES ARE SYNTHESIZED ON THE PHAGE DNA. *Molecular & General Genetics*, 1983. **191**(1): p. 22-25.
7. Rakonjac, J. and P. Model, Roles of pIII in filamentous phage assembly. *Journal of Molecular Biology*, 1998. **282**(1): p. 25-41.
8. Rakonjac, J., J.N. Feng, and P. Model, Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *Journal of Molecular Biology*, 1999. **289**(5): p. 1253-1265.
9. Baas, P.D., DNA-REPLICATION OF SINGLE-STRANDED ESCHERICHIA-COLI DNA PHAGES. *Biochimica Et Biophysica Acta*, 1985. **825**(2): p. 111-139.
10. Lerner, T.J. and P. Model, THE STEADY-STATE OF COLIPHAGE-F1 - DNA-SYNTHESIS LATE IN INFECTION. *Virology*, 1981. **115**(2): p. 282-294.
11. Pratt, D. and W.S. Erdahl, GENETIC CONTROL OF BACTERIOPHAGE M13 DNA SYNTHESIS. *Journal of Molecular Biology*, 1968. **37**(1): p. 181-&.
12. Mazur, B.J. and P. Model, REGULATION OF COLIPHAGE F1 SINGLE-STRANDED DNA-SYNTHESIS BY A DNA-BINDING PROTEIN. *Journal of Molecular Biology*, 1973. **78**(2): p. 285-300.
13. Forsheit, A.B., D.S. Ray, and L. Lica, REPLICATION OF BACTERIOPHAGE M13 .5. SINGLE-STRAND SYNTHESIS DURING M13 INFECTION. *Journal of Molecular Biology*, 1971. **57**(1): p. 117-&.
14. Fulford, W. and P. Model, BACTERIOPHAGE-F1 DNA-REPLICATION GENES .2. THE ROLES OF GENE-V PROTEIN AND GENE-II PROTEIN IN COMPLEMENTARY STRAND SYNTHESIS. *Journal of Molecular Biology*, 1988. **203**(1): p. 39-48.
15. Fulford, W. and P. Model, REGULATION OF BACTERIOPHAGE-F1 DNA-REPLICATION .1. NEW FUNCTIONS FOR GENE-II AND GENE-X. *Journal of Molecular Biology*, 1988. **203**(1): p. 49-62.
16. Kokoska, R.J. and D.A. Steege, Appropriate expression of filamentous phage f1 DNA replication genes II and X requires RNase E-dependent processing and separate mRNAs. *Journal of Bacteriology*, 1998. **180**(12): p. 3245-3249.
17. Steege, D.A., Emerging features of mRNA decay in bacteria. *Rna-a Publication of the Rna Society*, 2000. **6**(8): p. 1079-1090.
18. Stump, M.D. and D.A. Steege, Functional analysis of filamentous phage f1 mRNA processing sites. *Rna-a Publication of the Rna Society*, 1996. **2**(12): p. 1286-1294.

19. Yu, J.S., et al., In-frame overlapping genes: The challenges for regulating gene expression. *Molecular Microbiology*, 2007. **63**(4): p. 1158-1172.
20. Smits, M.A., et al., INITIATION AND TERMINATION SIGNALS FOR TRANSCRIPTION IN BACTERIOPHAGE-M13. *Nucleic Acids Research*, 1984. **12**(10): p. 4071-4081.
21. Edens, L., et al., MAPPING OF PROMOTER SITES ON GENOME OF BACTERIOPHAGE-M13. *European Journal of Biochemistry*, 1976. **70**(2): p. 577-587.
22. Blumer, K.J. and D.A. Steege, MESSENGER-RNA PROCESSING IN ESCHERICHIA-COLI - AN ACTIVITY ENCODED BY THE HOST PROCESSES BACTERIOPHAGE-F1 MESSENGER-RNAS. *Nucleic Acids Research*, 1984. **12**(4): p. 1847-1861.
23. Lafarina, M. and P. Model, TRANSCRIPTION IN BACTERIOPHAGE F1-INFECTED ESCHERICHIA-COLI MESSENGER POPULATIONS IN THE INFECTED CELL. *Journal of Molecular Biology*, 1983. **164**(3): p. 377-393.
24. Cashman, J.S., R.E. Webster, and D.A. Steege, TRANSCRIPTION OF BACTERIOPHAGE-FL - THE MAJOR INVIVO RNAS. *Journal of Biological Chemistry*, 1980. **255**(6): p. 2554-2562.
25. Blumer, K.J., M.R. Ivey, and D.A. Steege, TRANSLATIONAL CONTROL OF PHAGE-F1 GENE-EXPRESSION BY DIFFERENTIAL ACTIVITIES OF THE GENE V, VII, IX AND VIII INITIATION SITES. *Journal of Molecular Biology*, 1987. **197**(3): p. 439-451.
26. *Escherichia coli* and *Salmonella* cellular and molecular biology, in *Escherichia coli* and *Salmonella typhimurium*, F.C. Neidhardt and R. Curtiss, Editors. 1999, ASM Press: Washington, D.C.
27. Edens, L., R.N.H. Konings, and J.G.G. Schoenmakers, TRANSCRIPTION OF BACTERIOPHAGE M13 DNA - EXISTENCE OF PROMOTERS DIRECTLY PRECEDING GENES III, VI, AND I. *Journal of Virology*, 1978. **28**(3): p. 835-842.
28. Salis, H.M., THE RIBOSOME BINDING SITE CALCULATOR, in *Synthetic Biology*, Pt B: Computer Aided Design and DNA Assembly, C. Voigt, Editor 2011, Elsevier Academic Press Inc: San Diego. p. 19-42.
29. Salis, H.M., E.A. Mirsky, and C.A. Voigt, Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 2009. **27**(10): p. 946-U112.
30. Zaman, G., et al., REGULATION OF EXPRESSION OF THE GENOME OF BACTERIOPHAGE-M13 - GENE-V PROTEIN REGULATED TRANSLATION OF THE MESSENGER-RNAS ENCODED BY GENE-I, GENE-III, GENE-V AND GENE-X. *Biochimica Et Biophysica Acta*, 1991. **1089**(2): p. 183-192.
31. Zaman, G.J.R., J.G.G. Schoenmakers, and R.N.H. Konings, TRANSLATIONAL REGULATION OF M13 GENE-II PROTEIN BY ITS COGNATE SINGLE-STRANDED-DNA BINDING-PROTEIN. *European Journal of Biochemistry*, 1990. **189**(1): p. 119-124.
32. Opalka, N., et al., Structure of the filamentous phage pIV multimer by cryo-electron microscopy. *Journal of Molecular Biology*, 2003. **325**(3): p. 461-470.
33. Summers, W.C., *Phage and the Early Development of Molecular Biology in The Bacteriophages R. Calendar*, Editor 2006, Oxford University Press: Oxford ;.
34. Kutter, E., et al., *Phage Therapy in Clinical Practice: Treatment of Human Infections. Current Pharmaceutical Biotechnology*, 2010. **11**(1): p. 69-86.

35. Garcia, P., et al., Bacteriophages and their application in food safety. *Letters in Applied Microbiology*, 2008. **47**(6): p. 479-485.
36. Mahony, J., et al., Bacteriophages as biocontrol agents of food pathogens. *Current Opinion in Biotechnology*, 2011. **22**(2): p. 157-163.
37. Smith, G.P., FILAMENTOUS FUSION PHAGE - NOVEL EXPRESSION VECTORS THAT DISPLAY CLONED ANTIGENS ON THE VIRION SURFACE. *Science*, 1985. **228**(4705): p. 1315-1317.
38. Kehoe, J.W. and B.K. Kay, Filamentous phage display in the new millennium. *Chemical Reviews*, 2005. **105**(11): p. 4056-4072.
39. Flynn, C.E., et al., Viruses as vehicles for growth, organization and assembly of materials. *Acta Materialia*, 2003. **51**(19): p. 5867-5880.
40. Mao, C.B., A.H. Liu, and B.R. Cao, Virus-Based Chemical and Biological Sensing. *Angewandte Chemie-International Edition*, 2009. **48**(37): p. 6790-6810.
41. Hamzeh-Mivehroud, M., et al., Phage display as a technology delivering on the promise of peptide drug discovery. *Drug Discovery Today*, 2013. **18**(23-24): p. 1144-1157.
42. Ghosh, D., et al., M13-templated magnetic nanoparticles for targeted in vivo imaging of prostate cancer. *Nature Nanotechnology*, 2012. **7**(10): p. 677-682.
43. Nam, K.T., et al., Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science*, 2006. **312**(5775): p. 885-888.
44. Ge, H., A.J.M. Walhout, and M. Vidal, Integrating 'omic' information: a bridge between genomics and systems biology. *Trends in Genetics*, 2003. **19**(10): p. 551-560.
45. You, L.C., Toward computational systems biology. *Cell Biochemistry and Biophysics*, 2004. **40**(2): p. 167-184.
46. Thomas, R., BOOLEAN FORMALIZATION OF GENETIC-CONTROL CIRCUITS. *Journal of Theoretical Biology*, 1973. **42**(3): p. 563-585.
47. Clarke, B.L., STOICHIOMETRIC NETWORK ANALYSIS. *Cell Biophysics*, 1988. **12**: p. 237-253.
48. Endy, D., D. Kong, and J. Yin, Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7. *Biotechnology and Bioengineering*, 1997. **55**(2): p. 375-389.
49. Endy, D. and J. Yin, Toward antiviral strategies that resist viral escape. *Antimicrobial Agents and Chemotherapy*, 2000. **44**(4): p. 1097-1099.
50. Endy, D., et al., Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 2000. **97**(10): p. 5375-5380.
51. You, L. and J. Yin, Evolutionary design on a budget: robustness and optimality of bacteriophage T7. *Iee Proceedings Systems Biology*, 2006. **153**(2): p. 46-52.
52. You, L.C. and J. Yin, Patterns of Regulation from mRNA and Protein Time Series. *Metabolic Engineering*, 2000. **2**(3): p. 210-217.
53. You, L.C., P.F. Suthers, and J. Yin, Effects of Escherichia coli physiology on growth of phage T7 in vivo and in silico. *Journal of Bacteriology*, 2002. **184**(7): p. 1888-1894.
54. Chan, L.Y., S. Kosuri, and D. Endy, Refactoring bacteriophage T7. *Molecular Systems Biology*, 2005. **1**.
55. Hensel, S.C., J.B. Rawlings, and J. Yin, Stochastic Kinetic Modeling of Vesicular Stomatitis Virus Intracellular Growth. *Bulletin of Mathematical Biology*, 2009. **71**(7): p. 1671-1692.

56. Lim, K.I. and J. Yin, Computational Fitness Landscape for All Gene-Order Permutations of an RNA Virus. *Plos Computational Biology*, 2009. **5**(2).
57. Henry, M. and L. Debarbieux, Tools from viruses: Bacteriophage successes and beyond. *Virology*, 2012. **434**(2): p. 151-161.
58. Model, P., et al., THE REPLICATION OF BACTERIOPHAGE-F1 - GENE-V PROTEIN REGULATES THE SYNTHESIS OF GENE-II PROTEIN. *Cell*, 1982. **29**(2): p. 329-335.
59. Alberts, B., L. Frey, and H. Delius, ISOLATION AND CHARACTERIZATION OF GENE 5 PROTEIN OF FILAMENTOUS BACTERIAL VIRUSES. *Journal of Molecular Biology*, 1972. **68**(1): p. 139-&.
60. Kelman, Z. and M. Odonnell, DNA-POLYMERASE-III HOLOENZYME - STRUCTURE AND FUNCTION OF A CHROMOSOMAL REPLICATING MACHINE. *Annual Review of Biochemistry*, 1995. **64**: p. 171-200.
61. Asano, S., A. Higashitani, and K. Horiuchi, Filamentous phage replication initiator protein gpII forms a covalent complex with the 5 ' end of the nick it introduced. *Nucleic Acids Research*, 1999. **27**(8): p. 1882-1889.
62. Yen, T.S.B. and R.E. Webster, TRANSLATIONAL CONTROL OF BACTERIOPHAGE-F1 GENE-II AND GENE-X PROTEINS BY GENE-V PROTEIN. *Cell*, 1982. **29**(2): p. 337-345.
63. Fulford, W. and P. Model, GENE-X OF BACTERIOPHAGE-F1 IS REQUIRED FOR PHAGE DNA-SYNTHESIS - MUTAGENESIS OF IN-FRAME OVERLAPPING GENES. *Journal of Molecular Biology*, 1984. **178**(2): p. 137-153.
64. Goutelle, S., et al., The Hill equation: a review of its capabilities in pharmacological modelling. *Fundamental & Clinical Pharmacology*, 2008. **22**(6): p. 633-648.
65. Henry, T.J. and D. Pratt, PROTEINS OF BACTERIOPHAGE M13. *Proceedings of the National Academy of Sciences of the United States of America*, 1969. **62**(3): p. 800-&.
66. Pratt, D., P. Laws, and J. Griffith, COMPLEX OF BACTERIOPHAGE M13 SINGLE-STRANDED-DNA AND GENE-5 PROTEIN. *Journal of Molecular Biology*, 1974. **82**(4): p. 425-&.
67. Shimamoto, N. and H. Utiyama, MECHANISM AND ROLE OF COOPERATIVE BINDING OF BACTERIOPHAGE-FD GENE-5 PROTEIN TO SINGLE-STRANDED DEOXYRIBONUCLEIC-ACID. *Biochemistry*, 1983. **22**(25): p. 5869-5878.
68. Alma, N.C.M., et al., FLUORESCENCE STUDIES OF THE COMPLEX-FORMATION BETWEEN THE GENE-5 PROTEIN OF BACTERIOPHAGE-M13 AND POLYNUCLEOTIDES. *Journal of Molecular Biology*, 1983. **163**(1): p. 47-62.
69. Seeburg, P.H., C. Nusslein, and H. Schaller, INTERACTION OF RNA-POLYMERASE WITH PROMOTERS FROM BACTERIOPHAGE-FD. *European Journal of Biochemistry*, 1977. **74**(1): p. 107-113.
70. Seeburg, P.H. and H. Schaller, MAPPING AND CHARACTERIZATION OF PROMOTERS IN BACTERIOPHAGES FD, F1 AND M13. *Journal of Molecular Biology*, 1975. **92**(2): p. 261-&.
71. Kubitschek, H.E., CELL-VOLUME INCREASE IN ESCHERICHIA-COLI AFTER SHIFTS TO RICHER MEDIA. *Journal of Bacteriology*, 1990. **172**(1): p. 94-101.
72. Bremer, H. and P.P. Dennis, Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate in Escherichia coli and Salmonella cellular and molecular biology, F.C. Neidhardt and R. Curtiss, Editors. 1996, ASM Press: Washington, D.C.

73. Dennis, P.P. and H. Bremer, REGULATION OF RIBONUCLEIC-ACID SYNTHESIS IN ESCHERICHIA-COLI B/R - ANALYSIS OF A SHIFT-UP .1. RIBOSOMAL-RNA CHAIN GROWTH-RATES. *Journal of Molecular Biology*, 1973. **75**(1): p. 145-159.
74. Vogel, U. and K.F. Jensen, THE RNA CHAIN ELONGATION RATE IN ESCHERICHIA-COLI DEPENDS ON THE GROWTH-RATE. *Journal of Bacteriology*, 1994. **176**(10): p. 2807-2813.
75. Dennis, P.P., M. Ehrenberg, and H. Bremer, Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. *Microbiology and Molecular Biology Reviews*, 2004. **68**(4): p. 639-+.
76. Kushner, S.R., mRNA Decay in *Escherichia coli* and *Salmonella* cellular and molecular biology, F.C. Neidhardt and R. Curtiss, Editors. 1996, ASM Press: Washington, D.C.
77. Goodrich, A.F. and D.A. Steege, Roles of polyadenylation and nucleolytic cleavage in the filamentous phage mRNA processing and decay pathways in *Escherichia coli*. *Rna-a Publication of the Rna Society*, 1999. **5**(7): p. 972-985.
78. Zaman, G., J. Schoenmakers, and R. Konings, TRANSLATIONAL REGULATION IN BACTERIOPHAGE M13 - APPLICATION FOR THE REGULATED PRODUCTION OF HETEROLOGOUS PROTEINS IN ESCHERICHIA-COLI. 5th European Congress on Biotechnology, Proceedings, Vols 1 and 2, ed. C. Christiansen, L. Munck, and J. Villadsen 1990. 465-467.
79. Dennis, P.P. and H. Bremer, DIFFERENTIAL RATE OF RIBOSOMAL-PROTEIN SYNTHESIS IN ESCHERICHIA-COLI B-R. *Journal of Molecular Biology*, 1974. **84**(3): p. 407-422.
80. Iveyhoyle, M. and D.A. Steege, TRANSLATION OF PHAGE-F1 GENE-VII OCCURS FROM AN INHERENTLY DEFECTIVE INITIATION SITE MADE FUNCTIONAL BY COUPLING. *Journal of Molecular Biology*, 1989. **208**(2): p. 233-244.
81. Ploss, M. and A. Kuhn, Kinetics of filamentous phage assembly. *Physical Biology*, 2010. **7**(4).
82. Russel, M., FILAMENTOUS PHAGE ASSEMBLY. *Molecular Microbiology*, 1991. **5**(7): p. 1607-1613.
83. Kokoska, R.J., K.J. Blumer, and D.A. Steege, PHAGE-F1 MESSENGER-RNA PROCESSING IN ESCHERICHIA-COLI - SEARCH FOR THE UPSTREAM PRODUCTS OF ENDONUCLEASE CLEAVAGE, REQUIREMENT FOR THE PRODUCT OF THE ALTERED MESSENGER-RNA STABILITY (AMS) LOCUS. *Biochimie*, 1990. **72**(11): p. 803-811.
84. Endemann, H. and P. Model, LOCATION OF FILAMENTOUS PHAGE MINOR COAT PROTEINS IN PHAGE AND IN INFECTED-CELLS. *Journal of Molecular Biology*, 1995. **250**(4): p. 496-506.
85. Yu, J.S., S. Madison-Antenucci, and D.A. Steege, Translation at higher than an optimal level interferes with coupling at an intercistronic junction. *Molecular Microbiology*, 2001. **42**(3): p. 821-834.
86. Rapoza, M.P. and R.E. Webster, THE PRODUCTS OF GENE-I AND THE OVERLAPPING IN-FRAME GENE-XI ARE REQUIRED FOR FILAMENTOUS PHAGE ASSEMBLY. *Journal of Molecular Biology*, 1995. **248**(3): p. 627-638.
87. Zaman, G.J.R., et al., GENE-V PROTEIN-MEDIATED TRANSLATIONAL REGULATION OF THE SYNTHESIS OF GENE-II PROTEIN OF THE

- FILAMENTOUS BACTERIOPHAGE-M13 - A DISPENSABLE FUNCTION OF THE FILAMENTOUS-PHAGE GENOME. *Journal of Bacteriology*, 1992. **174**(2): p. 595-600.
88. Zinder, N.D. and K. Horiuchi, MULTIREGULATORY ELEMENT OF FILAMENTOUS BACTERIOPHAGES. *Microbiological Reviews*, 1985. **49**(2): p. 101-106.
89. Merriam, V., STABILITY OF CARRIER STATE IN BACTERIOPHAGE-M13-INFECTED CELLS. *Journal of Virology*, 1977. **21**(3): p. 880-888.
90. Hoffmannberling, H. and R. Maze, RELEASE OF MALE-SPECIFIC BACTERIOPHAGES FROM SURVIVING HOST BACTERIA. *Virology*, 1964. **22**(3): p. 305-&.
91. Phage display a laboratory manual, ed. C.F. Barbas2001, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Appendix A – Simbiology Inputs

Table A1 – Contains each reaction and rate law equation which was implemented in the simbiology program.

Reaction Number	Reaction Equation	Rate Equation
Replication Equations		
1	$ssDNA + DP3 \rightarrow ssPDNA$	$Rate = C_1 * ssDNA * DP3$
2	$ssPDNA \rightarrow RF1 + DP3 + DA + DB + DH + DZ + DW$	$Rate = C_2 * ssPDNA$
3	$RF1 + P2 \leftrightarrow RF2$	$Rate = C_3 * RF1 * P2 - C_4 * RF2$
4	$RF2 + DP3 + DA + DB + DH + DZ + DW \rightarrow RF2DP3$	$Rate = C_1 * RF2 * DP3$
5	$RF2DP3 \rightarrow P2 + DP3 + RF1 + ssDNA + DA + DB + DH + DZ + DW$	$Rate = C_2 * RF2DP3$
6	$P2 + P10 \leftrightarrow P2P10$	$Rate = C_5 * P2 * P10 - C_6 * P2P10$
7	$ssDNA + 1600 P5 \rightarrow P5DNA$	$Rate = C_7 * \frac{P5^{n_1}}{k_{m1}^{n_1} + P5^{n_1}} ssDNA * P5$
Transcription Equations		
8	$DA + RNAP \rightarrow EA$	$Rate = C_{8,A} * DA * RNAP$
9	$DB + RNAP \rightarrow EB$	$Rate = C_{8,B} * DB * RNAP$
10	$DH + RNAP \rightarrow EH$	$Rate = C_{8,H} * DH * RNAP$
11	$DZ + RNAP \rightarrow EZ$	$Rate = C_{8,Z} * DZ * RNAP$
12	$DW + RNAP \rightarrow EW$	$Rate = C_{8,W} * DW * RNAP$
13	$EA \rightarrow DA + ELA$	$Rate = C_9 * EA$
14	$EB \rightarrow DB + ELB$	$Rate = C_9 * EB$
15	$EH \rightarrow DH + ELH$	$Rate = C_9 * EH$
16	$EZ \rightarrow DZ + ELZ$	$Rate = C_9 * EZ$
17	$EW \rightarrow DW + ELW$	$Rate = C_9 * EW$
18	$ELA \rightarrow RNAP + A + RBS2 + RBS5 + RBS7 + RBS9 + RBS8$	$Rate = C_{10,A} * ELA$
19	$ELB \rightarrow RNAP + B + RBS10 + RBS5 + RBS7 + RBS9 + RBS8$	$Rate = C_{10,B} * ELB$
20	$ELH \rightarrow RNAP + H + RBS9 + RBS8$	$Rate = C_{10,H} * ELH$

Reaction Number	Reaction Equation	Rate Equation
21	$ELZ \rightarrow RNAP + C_{11} Z + RBS3 + RBS6 + (1 - C_{11})(Y + RBS1 + RBS11)$	$Rate = C_{10,Z} * ELZ$
22	$ELW \rightarrow RNAP + W + RBS4$	$Rate = C_{10,W} * ELW$
<u>mRNA Degradation Equations</u>		
23	$A + RBS2 \rightarrow D$	$Rate = C_{12,A} * [A]$
24	$B + RBS10 \rightarrow D$	$Rate = C_{12,B} * [B]$
25	$D \rightarrow E$	$Rate = 0.7 C_{12,D} * [D]$
26	$E \rightarrow F$	$Rate = 0.7 C_{12,E} * [E]$
27	$F + RBS5 \rightarrow G$	$Rate = 0.7 C_{12,F} * [F]$
28	$G \rightarrow H$	$Rate = 0.7 C_{12,G} * [G]$
29	$H + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = C_{12,H} * [H]$
30	$D + RBS5 + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,D} * [D]$
31	$E + RBS5 + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,E} * [E]$
32	$F + RBS5 + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,F} * [F]$
33	$G + RBS9 + RBS8 \rightarrow \emptyset$	$Rate = 0.3 C_{12,G} * [G]$
34	$W + RBS4 \rightarrow \emptyset$	$Rate = C_{12,W} * [W]$
35	$Y + RBS6 + RBS3 + RBS11 + RBS1 \rightarrow \emptyset$	$Rate = C_{12,Y} * [Y]$
36	$Z + RBS6 + RBS3 \rightarrow \emptyset$	$Rate = C_{12,Z} * [Z]$
<u>Translation Equations</u>		
37	$RBS2 + R \rightarrow RBS2R$	$Rate = C_{13,2} * RBS2 * R * \frac{1}{1 + (P^5/Ef2 * (RBS2 + 1))^n}$
38	$RBS10 + R \rightarrow RBS10R$	$Rate = C_{13,10} * RBS10 * R * \frac{1}{1 + (P^5/Ef10 * (RBS10 + 1))^n}$
39	$RBS5 + R \rightarrow RBS5R$	$Rate = C_{13,5} * RBS5 * R * \frac{1}{1 + (P^5/Ef5 * (RBS5 + 1))^n}$

Reaction Number	Reaction Equation	Rate Equation
40	$RBS9 + R \rightarrow RBS9R$	$Rate = C_{13,9} * RBS9 * R$
41	$RBS8 + R \rightarrow RBS8R$	$Rate = C_{13,8} * RBS8 * R$
42	$RBS3 + R \rightarrow RBS3R$	$Rate = C_{13,3} * RBS3 * R * \frac{1}{1 + (P^5/Ef3 * (RBS3 + 1))^n}$
43	$RBS6 + R \rightarrow RBS6R$	$Rate = C_{13,6} * RBS6 * R$
44	$RBS1 + R \rightarrow RBS1R$	$Rate = C_{13,1} * RBS1 * R * \frac{1}{1 + (P^5/Ef1 * (RBS1 + 1))^n}$
45	$RBS11 + R \rightarrow RBS11R$	$Rate = C_{13,11} * RBS11 * R$
46	$RBS4 + R \rightarrow RBS4R$	$Rate = C_{13,4} * RBS4 * R$
47	$RBS2R \rightarrow RBS2 + PD2$	$Rate = C_{14} * RBS2R$
48	$RBS10R \rightarrow RBS10 + PD10$	$Rate = C_{14} * RBS10R$
49	$RBS5R \rightarrow RBS5 + PD5$	$Rate = C_{14} * RBS5R$
50	$RBS9R \rightarrow RBS9 + R + P7$	$Rate = C_{14} * RBS9R$
51	$RBS8R \rightarrow RBS8 + PD8$	$Rate = C_{14} * RBS8R$
52	$RBS3R \rightarrow RBS3 + PD3$	$Rate = C_{14} * RBS3R$
53	$RBS6R \rightarrow RBS6 + PD6$	$Rate = C_{14} * RBS6R$
54	$RBS1R \rightarrow RBS1 + PD1$	$Rate = C_{14} * RBS1R$
55	$RBS11R \rightarrow RBS11 + P11$	$Rate = C_{14} * RBS11R$
56	$RBS4R \rightarrow RBS4 + PD4$	$Rate = C_{14} * RBS4R$
57	$PD2 \rightarrow P2 + R$	$Rate = C_{15,2} * PD2$
58	$PD10 \rightarrow P10 + R$	$Rate = C_{15,10} * PD10$
59	$PD5 \rightarrow P5 + 0.1 P7 + R$	$Rate = C_{15,5} * PD5$
60	$PD8 \rightarrow P8 + R$	$Rate = C_{15,8} * PD8$
61	$PD3 \rightarrow P3 + R$	$Rate = C_{15,3} * PD3$
62	$PD6 \rightarrow P6 + R$	$Rate = C_{15,6} * PD6$
63	$PD1 \rightarrow P1 + R$	$Rate = C_{15,1} * PD1$
64	$PD11 \rightarrow P11 + R$	$Rate = C_{15,11} * PD11$
65	$PD4 \rightarrow P4 + R$	$Rate = C_{15,4} * PD4$

Reaction Number	Reaction Equation	Rate Equation
<u>Phage Assembly Equations</u>		
66	$14 P1 + 14 P4 + 14 P11 \rightarrow A_S$	$Rate = C_{16} * P1 * P4 * P11$
67	$A_S + 5 P7 + 5 P9 \rightarrow P_I$	$Rate = C_{17} * A_S * P7 * P9$
68	$P_I + P5DNA \rightarrow P_E$	$Rate = C_{18} * P_I * P5DNA$
69	$P_E + 2700 P8 \rightarrow P_F + 1600 P5$	$Rate = C_{19} * [P_E] * \frac{P8^{n_2}}{K_{m1}^{n_2} + P8^{n_2}}$
70	$P_F + 5 P3 + 5 P9 \rightarrow Phage + A_S$	$Rate = C_{20} * P3 * P9$

Table A2 – Contains a description, numerical value, and reference for each reaction constant. N/A indicates the reaction constant was not found through literature and the value is fitted to achieve desired results. # - indicates the relative ratios of the two parameters were estimated but the magnitude of the value is unknown. Parameters with reference of AVE indicate parameters which were taken as the average value between the high and low values of the measured half-lives of the more frequently transcribed region. ### - indicates the relative ribosome binding rates were known but the rate of binding was unknown.

Parameter Description	Symbol	Value	Reference
Replication			
DNA Polymerase III Binding Rate to ssDNA and dsDNA	C_1	$0.1 \text{ (Molecule*Second)}^{-1}$	N/A
DNA Polymerase III Elongation Constant	C_2	$0.117 \text{ (Second)}^{-1}$	[60]
P2 Nicking RF1 DNA to Form RF2 DNA	C_3	$1 \text{ (Molecule*Second)}^{-1}$	[61] [#]
P2 Dissociation to RF2 DNA	C_4	$1.0E2 \text{ (Second)}^{-1}$	[61] [#]
P2/P10 Association Constant	C_5	$1 \text{ (Molecule*Second)}^{-1}$	N/A
P2/P10 Dissociation Constant	C_6	$10^{-15} \text{ (Second)}^{-1}$	N/A
Rate of P5 Sequestering ssDNA	C_7	$1E-3 \text{ (Molecule*Second)}^{-1}$	N/A
Threshold of P5 Molecules for P5DNA Formation	K_{m1}	1100 (Molecules)	[68]
Hill Coefficient for P5DNA Formation	n_1	1 (Unitless)	N/A
Transcription			
RNAP Binding to Promoter Site A	$C_{8,A}$	$6E-3 \text{ (Molecule*Second)}^{-1}$	[21, 69]
RNAP Binding to Promoter Site B	$C_{8,B}$	$1E-2 \text{ (Molecule*Second)}^{-1}$	[21, 69]
RNAP Binding to Promoter Site H	$C_{8,H}$	$6.5E-2 \text{ (Molecule*Second)}^{-1}$	[21, 69]
RNAP Binding to Promoter Site W	$C_{8,W}$	$2E-3 \text{ (Molecule*Second)}^{-1}$	N/A
RNAP Binding to Promoter Site Z	$C_{8,Z}$	$2E-4 \text{ (Molecule*Second)}^{-1}$	N/A
RNAP Clearing the Promoter Site	C_9	$1.34 \text{ (Second)}^{-1}$	[73-75]
RNAP Elongation of mRNA A	$C_{10,A}$	$3.6E-2 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA B	$C_{10,B}$	$7.2E-2 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA H	$C_{10,H}$	$2.5E-1 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA W	$C_{10,W}$	$5.5E-2 \text{ (Second)}^{-1}$	[73, 74]
RNAP Elongation of mRNA Z	$C_{10,Z}$	$4.3E-2 \text{ (Second)}^{-1}$	[73, 74]
Efficiency of Rho-Dependent Terminator Distal to P3	C_{11}	0.60 (Unitless)	[5]
Degradation			
Degradation Rate of mRNA A	$C_{12,A}$	$2.0E-2 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA B	$C_{12,B}$	$5.8E-2 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA D	$C_{12,D}$	$5.8E-2 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA E	$C_{12,E}$	$4.6E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA F	$C_{12,F}$	$1.9E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA H	$C_{12,G}$	$1.4E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA G	$C_{12,H}$	$1.2E-3 \text{ (Second)}^{-1}$	[24]
Degradation Rate of mRNA W	$C_{12,W}$	$1.9E-3 \text{ (Second)}^{-1}$	AVE
Degradation Rate of mRNA Y	$C_{12,Y}$	$1.9E-3 \text{ (Second)}^{-1}$	AVE
Degradation Rate of mRNA Z	$C_{12,Z}$	$1.9E-3 \text{ (Second)}^{-1}$	AVE
Translation			
Ribosome Binding to P1 Start Codon	$C_{13,1}$	$1.60 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P2 Start Codon	$C_{13,2}$	$10.8 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P3 Start Codon	$C_{13,3}$	$8.00 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P4 Start Codon	$C_{13,4}$	$4.60 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P5 Start Codon	$C_{13,5}$	$20.0 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P6 Start Codon	$C_{13,6}$	$0.80 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P8 Start Codon	$C_{13,8}$	$4.40 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P9 Start Codon	$C_{13,9}$	$1.60 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Binding to P10 Start Codon	$C_{13,10}$	$1.08 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}

Parameter Description	Symbol	Value	Reference
<u>Translation (Continued)</u>			
Ribosome Binding to P11 Start Codon	$C_{13,11}$	$4.00 \text{ (Molecule*Second)}^{-1}$	[28, 29] ^{##}
Ribosome Clearing Start Codon	C_{14}	$0.25 \text{ (Second)}^{-1}$	[48, 79]
Ribosome Elongation of P1	$C_{15,1}$	$4.8\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P2	$C_{15,2}$	$4.1\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P3	$C_{15,3}$	$3.7\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P4	$C_{15,4}$	$3.8\text{E-}2 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P5	$C_{15,5}$	$4.3\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P6	$C_{15,6}$	$2.5\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P8	$C_{15,8}$	$8.9\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P10	$C_{15,10}$	$2.4\text{E-}1 \text{ (Second)}^{-1}$	[79]
Ribosome Elongation of P11	$C_{15,11}$	$3.1\text{E-}1 \text{ (Second)}^{-1}$	[79]
Efficiency of P5 inhibiting P1 Translation	EF1	80 (Unitless)	N/A
Efficiency of P5 inhibiting P2 Translation	EF2	$1.5\text{E-}4$ (Unitless)	N/A
Efficiency of P5 inhibiting P3 Translation	EF3	4 (Unitless)	N/A
Efficiency of P5 inhibiting P5 Translation	EF5	$1.0\text{E-}2$ (Unitless)	N/A
Efficiency of P5 inhibiting P10 Translation	EF10	$2.6\text{E-}5$ (Unitless)	N/A
Hill Coefficient for P5 Inhibiting Translation	n_2	1 (Unitless)	N/A
<u>Assembly</u>			
Assembly Site Formation	C_{16}	$0.1 \text{ (Molecule}^2\text{*Second)}^{-1}$	N/A
P_I Formation	C_{17}	$0.1 \text{ (Molecule*Second)}^{-1}$	N/A
P_E Formation	C_{18}	$0.1 \text{ (Molecule*Second)}^{-1}$	N/A
Phage Elongation/P8 Addition	C_{19}	$1.2\text{E-}3 \text{ (Second)}^{-1}$	[7, 82]
Phage Detaching from Cell Membrane	C_{20}	$0.1 \text{ (Molecule}^2\text{*Second)}^{-1}$	N/A
Number of P8 Molecules Needed for Elongation to Occur	K_{m2}	3000 (Molecules)	N/A
Hill Coefficient	n_3	40 (Unitless)	N/A

Appendix B – Simulation Strategy

Introduction to the Tools Used to Develop the Model

The model was developed in Matlab with the aid of an additional program Simbiology (a complementary program which runs with Matlab). Simbiology is a graphical-user interface where the model equations and rate constants can be stored. It is a useful tool because it can convert reaction equations into mass action kinetic laws or any other rate laws. For example, simbiology can convert the reaction $A + B \rightarrow C$ into a mass action kinetic equation. Once you have inserted the equation, the second order rate constant and the initial value of all species at time zero must be defined in order to simulate the reaction. Insert the equation, rate law, and kinetic law and the program will convert it to a rate law as seen below:

$$Rate = k * A * B$$

If you choose to run Simbiology it will automatically develop three sets of ordinary differential equations, see below, and solve it using a numerical method. Multiple numerical methods exist and you must choose one to fit your model. See the Simbiology user manual for more information on the different numerical methods.

$$\frac{dA}{dt} = -k * A * B \quad \frac{dB}{dt} = -k * A * B \quad \frac{dC}{dt} = k * A * B$$

In the creation of the M13 phage infection model Simbiology greatly simplified the process of modifying the model. For example, what if we find out the reaction between A and B is also reversible. In Simbiology we would just need to update the reaction to read $A + B \leftrightarrow C$ and provide a first order rate constant (k_r). In return, Simbiology would generate an updated group of differential equations:

$$\frac{dA}{dt} = -k * A * B + k_r * C \quad || \quad \frac{dB}{dt} = -k * A * B + k_r * C \quad || \quad \frac{dC}{dt} = k * A * B - k_r * C$$

The simple act of adding a reverse reaction has caused all three differential equations to change. As the models become more complex, simple changes to the reactions can cause many changes to many differential equations. Simbiology significantly reduces this complexity for you by developing the equations each time. I would direct the reader to the help resources provided by the MathWorks website for more information on Simbiology.

Although the model can be simulated directly from the Simbiology interface, additional files were created in m-files. m-files is a document which contains a list of Matlab commands, such as for loops, while loops, and etc. Once again I would refer the reader to the MathWorks website or the Matlab help menu for more information. Three separate were created and are necessary to run a successful simulation of phage model. Their functions included: opening the Simbiology file containing the model, updating all parameters, running the simulation, and plotting species vs. time graphs. It should be noted, the simulation could be run directly from the Simbiology interface, however it was easier to use m-files to update and simulate the model.

Files Needed for the Analysis of M13 Infection

In order to run the model successfully you must have a computer with Matlab installed on it with the additional program Simbiology. Simbiology is a separate program which must be purchased additionally. To check if your computer has Simbiology simply type in simbiology (in all lower case letters) into the command window and hit enter. If you do not have Simbiology installed you will get an error telling you simbiology is not defined.

Once you have Matlab and Simbiology successfully installed on your computer, the M13 phage biology model can be run from your computer if you have all four of the following files in the same folder on your computer:

1. ModelRun.m
2. Load_Parameters.m
3. Plot_Phage_Species.m
4. Final_Model.sbroj

The (.m) files indicate simple Matlab files and the (.sbroj) indicates the simbiology file. The previous file can be obtained by asking Dr. Nick Fisk. To run the simulation open the ModelRun.m files in Matlab and press the play button on the top of the screen. The model will simulate and produce species vs. time graphs depending on the options chosen in the files. Two additional files were created which plot special graphs used in the formation of the paper:

5. PlotsForPaper.m
6. HostResourceUtilization.m

The “PlotsForPaper.m” file contains the information to generate all of the graphs displayed in the “Validation of the Model” section in Chapter 3 of the thesis. It should be noted, more graphs are generated in the file than appeared in the thesis. The “HostResourceUtilization.m” file will generate all graphs in the “Utilization of Host Resource Section.”

The files were separated to divide the information into easily understandable sub-blocks. It is important to note the two files named Load_Parameters.m and Plot_Phage_Species.m are a special type of m-file called function files. Function files require input from either the user or another m-file in order to run. In other words, these files will not function by themselves.

The plots created in the “Exploring the Effects of Removing P5 Inhibition of Proteins 1, 2, 3, 5, 10” section were generated using two m-files:

7. RemovingP5Inhibition.m
8. Comparision.m

The RemovingP5Inhibition.m file is identical to the ModelRun.m file except it takes in an extra input called WhichOneToInhibit. WhichOneToInhibt tells the computer to turn off the translational inhibition effects of P5 on a specific protein. It should be noted that one protein can be removed from P5’s negative effects or all of the proteins. Secondly, the simulation is run twice in the file, the first time is identical to the ModelRun.m simulation and the second time turns off the negative effects of P5 inhibition of the proteins you chose. Comparison.m file is run after the RemovingP5Inhibition.m file is ran and it generates the plots in the thesis paper.

In order to simulate cellular replication as described in section 3.5 in the thesis three additional files were created however they were modifications to already existing files ModelRun.m, Plot_Phage_Species.m, and PlotsForPaper.m.

9. ReplicationAddition.m
10. Replication_PPS.m
11. PlotReplicationPaper.m

The ReplicationAddition.m m-file is identical to the ModelRun.m m-file however it has a couple modifications. First it takes in a new input called No_Of_Replications which is the number of cell divisions the model will simulate. If No_Of_Replications = 0 then the m-file will run identical to the ModelRun.m m-file. Secondly there is code to re-start the simulation with half of the final values from the previous simulation to be the initial values of the new

simulation. Last, the time is converted from seconds to number of replications. Replication_PPS.m and PlotReplicationPaper.m files are identical to the Plot_Phage_Species.m, and PlotsForPaper.m except for the x-axis is labeled 'No Of Replications'.

The final two files which were used for results and discussions in the paper are:

12. ChangingAssemblyElongation.m

13. ChangingAssemblyElongationParamtersChanged.m

The first file decreases the rate of phage assembly and was used to analyze how the other model species would react. The second file does the same thing but more parameters were changed to try to get the most of the same output from most species as seen the "Validation of the Mode" section.

General Strategy in the Creation of the m-Files

The "ModelRun.m" m-file is the central file for running the M13 phage infection simulation. It is also basically identical to four other files discussed in the previous paragraphs (RemovingP5Inhibition.m, ReplicationAddition.m, ChangingAssemblyElongation.m, and ChangingAssemblyElongationParamtersChanged.m). The following will list important features of the file which may not be obvious from the comments.

1. ModelRun.m

- a. The main file which is responsible for calling up the other m-files necessary for a successful simulation. This is the file where you hit the run button on top of the screen.
- b. Requires 3 additional files in order for it to run properly:
 - i. The simbiology file containing the model "Final_Model.sbproj"

- ii. The user-defined function file for Load_Parameter.m function file
 - iii. The user-defined function file for Plot_Phage_Species.m file
- c. Inputs Include:
 - i. What to plot section has 12 inputs where each input is category you of species you may plot at the end of the simulation. If you change the first input which is described as “All DNA species will be plotted” from a 0 to a 1, this will tell the computer to plot all DNA species. It also applies to the rest of the variables in the section.
- d. Load_Model
 - i. This is the file name of the simbiology file that has the model information stored into it. By default the name is called “ModelForPaper.sbproj”
- e. The Simulation_Time input tells the computer for how many seconds in the phage infection you wish to simulate.
- f. The next two inputs are P and PC
 - i. P= pre-allocating a vector of zeros where the simulation parameters will be stored into.
 - ii. PC = pre-allocating a vector of zeros where the inputs must be converted to rate constants in the Load_Parameters.m function
- g. The next set of parameters is the replication inputs (1-7). The values displayed on this page are the values used for the appropriate rate constant.
- h. PC(1) is the first parameter that must be converted from rate of DNA polymerase elongation (Nucleotides per second) to a rate constant (1/second)

- i. Same story as it goes down, if the parameter is inputted to the P vector it is used directly, if it is placed into the PC vector it must be calculated
- j. Initial_Amount
 - i. This is the vector where you can change the initial amount of host resources which include DNA polymerase, Ribosomes, RNA polymerase
- k. sbioloadproject
 - i. This will tell MATLAB to access the model assigned to Load_Model variable
- l. m1.Configset.StopTime tells the computer for how long to run the simulation for (simulation time NOT computer time)
- m. The Load_Parameter(PC) function (which is user defined) takes the input of parameters and turns them into usable rate constants. The output is stored to the appropriate location in the P vector, which is the designation for the exact parameters used in the simulation
- n. Lines 120-122 update the parameters in the model from the stored vector.
 - i. **!!!!!!!!!!!!!!!!!!!!Word of caution for future users!!!!!!!!!!!!!!!!!!!!**
 - ii. **The order of which the P vector values are inputted is very important. The first space (index = 1) in the P vector is the rate constant for P2 nicking dsDNA. This corresponds to the first parameter inputted into the simbiology file under the parameter index. Those two parameters must match in order to use the simulation correctly. They are matched now, but if you delete any**

parameter value in simbiology, make sure to double check they match.

- o. Lines 124-126 initial amount of host resources
 - i. Updates the initial value of the host resources from the user input.
 - ii. The species index in the m-file must match the index in simbiology
 - p. Sbiosimulate will simulate the model storing the time vector in variable t and species matrix in variable x
 - i. The matrix x is set up where each column represents a species, and each row represents a time step
 - q. I converted time from seconds to minutes in line 142
 - r. Plot_Phage_Species is an user defined function where the selected categories described above are calculated to be plotted are then plotted
2. Load_Parameters.m
- a. User defined function that takes in a vector of inputs described in the file itself.
 - b. The information contained in this file includes length of genome (in nucleotides), length of each gene within genome, relative promoter strengths and ribosomes binding site affinities
 - c. The output include rate constants for the rate of DP3 elongation, rate of RNAP binding to all promoters, rate of RNAP elongation of all species, rate of ribosomes binding to all binding sites, rate of ribosomes elongation, and rate of mRNA degradation
3. Plot_Phage_Species
- a. User defined function that takes three inputs in

- i. A vector of zeros and ones which will decide which groups of species will be plotted
- ii. The time vector and species matrix generated from the simulation
- iii. Output are species vs. time plots

4. Final_Model.sbproj

- a. This is the simbiology file responsible for storing differential equations
- b. One word of caution that is the same as before but I must reiterate!
 - i. If you add a parameter or species within the corresponding lists, make sure the indices still correspond to the indices in the m-files. For example, lets say you add a new DNA species to the model and place it in the 6th index within simbiology. Running the m-files will label each graph incorrectly. The plot for the P5DNA vs. time graph on the new simulation will actually be the new DNA species vs. time graph.

5. PlotsForPaper.m

- a. Just like the parameters, the species are index in simbiology and take on an integer value. For example, RF1 takes on an index value of 3. When PlotsForPaper.m plots RF1 it is found by inputted the index of 3 to plot the values. It should be noted that the simbiology index and m-file indexes are identical and if you delete an index in simbiology you must update the index in the m-files.
 - i. Actually, this is a bad way of doing things, maybe try to think of a better way to program this into Matlab so changes in simbiology reflect changes in m-files automatically.

- b. Important to note: this file will only run when the ModelRun.m file has already completed the simulation. Do not type in `clc` or `clear all` in the command line before running it because it will remove the required information needed to plot the diagrams.
 - c. This file was created to generate a set of plots that are easily combined for the paper. It is different from the `Plot_Phage_Species.m` file because that plots 1 species per plot
 - d. This file combines related species for easy graphical explanation on the paper
 - e. All plots talked about in the Validation of the model section can be seen here
6. `HostResourceUtilization.m`
- a. Important to note: this file will only run when the ModelRun.m file has already completed the simulation. Do not type in `clc` or `clear all` in the command line before running it because it will remove the required information needed to plot the diagrams
 - b. The file will generate all of the graphs used for the how resources are utilized section
 - c. Once again, this file must be run AFTER the ModelRun.m has ran and stored the simulation results

Appendix C: Source Code

Appendix C contains the three m-files created in order to run the simulation which include the:

1. ModelRun.m
2. Load_Parameters.m
3. Plot_Phage_Species.m

ModelRun.m

```
%-----  
%Model Run-----Version 3-----February 27,2014  
%-----  
clc; clear all  
  
%-----  
%-----What to Plot-----  
%-----  
Plot_What = zeros(1,12);  
Plot_What(1) = 0; %All DNA species will be plotted  
Plot_What(2) = 0; %Free RNAP promoter sites  
Plot_What(3) = 0; %RNAP promoter sites bound to a RNA polymerase  
Plot_What(4) = 0; %Elongating RNAP away from promoter site  
Plot_What(5) = 0; %Polycistronic mRNA species  
Plot_What(6) = 0; %Free Ribosome Binding sites  
Plot_What(7) = 0; %Ribosome binding site with ribosome attached  
Plot_What(8) = 0; %Elongating Ribosomes away from ribosome binding site  
Plot_What(9) = 0; %Phage Proteins  
Plot_What(10) = 0; %New Phage  
Plot_What(11) = 0; %Assembly Site formation species  
Plot_What(12) = 0; %E.Coli Proteins  
%-----  
%-----  
%-----  
  
%-----  
%-----Inputs-----  
%-----  
  
%-----  
%-----General-----  
%-----Defining Which Model to Use .sproj file  
Load_Model = 'Final_Model.sbproj';  
%-----The length of Simulation Time in seconds  
Simulation_Time = 3600*1;  
%-----Parameters That May be Varied  
P = zeros(1,66); %Empty Vector that will contain all constant variables  
%-----Parameters That need to be converted to rate constants  
PC = zeros(1,16); %Empty Vector  
%-----
```



```

%-----Replication-----
P(1) = 1;      %Forward Reaction Rate for P2 nicking DNA (1/(molecule*second))
P(2) = P(1)*1000;      %Reverse Reaction Rate for P2 nicking DNA (1/second)
P(3) = 1;      %P2/P10 forward binding constant (1/(molecule*second))
P(4) = 9.5454e-16;      %P2/P10 reverse binding constant (1/(second))
P(5) = .001;      %Rate of P5 Sequestering ssDNA
P(6) = 1100;      %Number of P5 molecules needed to start S.S. DNA Production
P(7) = 0.1;      %Rate of DNA polymerase Binding
PC(1) = 800;      %Rate of DNA polymerase elongation (Nucleotides/S)
%-----

%-----Transcription-----
PC(2) = 0.01;      %RNA Polymerase Binding (1/(molecule*second))
PC(3) = 67;      %Rate Of RNAP elongation (Nucleotides/S)
%-----

%-----Translation-----
PC(4) = 20;      %Ribosome Polymerase Binding (1/(molecule*second))
PC(5) = 14*3;      %Ribosome Elongation Rate (nucleotides/second)
P(61) = 0.00015;      %Efficiency of P5 Inhibition to P2
P(62) = 0.000026;      %Efficiency of P5 Inhibition to P10
P(63) = 0.01;      %Efficiency of P5 Inhibition to P5
P(64) = 4;      %Efficiency of P5 Inhibition to P3
P(65) = 80;      %Efficiency of P5 Inhibition to P1
%Note-An efficiency factor less than 1, will increase the inhibition of P5
%on translation. An efficiency factor greater than one will decrease the
%effect P5 has on inhibition.
%-----

%-----mRNA Degradation-----
PC(6) = 3.4/6;      %Half-Life of mRNA A in minutes
PC(7) = 2;      %Half-Life of mRNA B in minutes
PC(8) = 2;      %Half-Life of mRNA C in minutes
PC(9) = 2;      %Half-Life of mRNA D in minutes
PC(10) = 2.5;      %Half-Life of mRNA E in minutes
PC(11) = 6;      %Half-Life of mRNA F in minutes
PC(12) = 8;      %Half-Life of mRNA G in minutes
PC(13) = 10;      %Half-Life of mRNA H in minutes
PC(14) = 6;      %Half-Life of mRNA Z in minutes
PC(15) = 6;      %Half-Life of mRNA Y in minutes
PC(16) = 6;      %Half-Life of mRNA W in minutes
%-----

%-----Phage Production-----

P(55) = 0.00123;      %Rate of new phage elongation (1/s)
P(56) = 0.1;      %Rate of assembly site formation (1/(Second*Molecule))
P(57) = 0.1;      %Rate of P7/P9 forming a complex with an assembly site
      %(1/(Second*Molecule^2))
P(58) = 0.1;      %Rate of P5DNA binding to assembly site containing p7
      %and P9 (1/(Second*Molecule))
P(59) = 0.1;      %Rate of Phage Termination (1/(Second*Molecule^2))

P(60) = 3000;      %Minimum number of P8 molecules needed to have phage
      %producing at full saturation (Molecule)

```

```

%-----

%-----Hill Coefficient Values-----
P(66) = 40; %This one makes sure no negative P8 is being made, that is why
it is so high
P(67) = 1; %This is the one that dictates the P5 inhibition reactions
P(68) = 1; %This the hill coefficient constant for the rate of P5DNA
formation
%-----

%-----Initial Amount of E.Coli Proteins-----
Initial_Amount = [
    3 %DNA Polymerase
    7880 %Ribosomes
    1280 %RNA Polymerases [1500-11400 Total]
];
%-----

%-----Loading Model and Parameters-----
%-----Loading the Model
sbioloadproject(Load_Model)
%-----Defining the Legnth of the Simulation Time
m1.Configset.StopTime = Simulation_Time;
%-----Calculation of the constants that are under the PC array
[P(8:54)] = Load_Parameters(PC);
%-----Loading the Parameters
for n = 1:length(P)
    m1.parameters(n).value = P(n);
end
%---Loading the Initial Amount of E.Coli Species
for n = 75:77
    m1.species(n).InitialAmount = Initial_Amount(n-74);
end
%-----

%-----Simulating The Model-----
%-----

[t,x] = sbiosimulate(m1);
%-----

%-----Plotting Species-----
%-----
t = t./60;

```

```
Plot_Phage_Species(Plot_What,t,x)
```

```
%-----
%-----
%-----

%-----Pulling Final Values-----
%-----
Total_P5DNA = x(end,6)+ x(end,80) + x(end,81)
DNA = x(end,3) + x(end,4) + x(end,5)
P5DNA = x(end,6) + x(end,80) +x(end,81)
P1 = x(end,70) + 14.*(x(end,78) + x(end,79) + x(end,80) + x(end,81))
P2 = x(end,62) + x(end,73)
P3 = x(end,68)
P4 = x(end,72) + 14.*(x(end,78) + x(end,79) + x(end,80) + x(end,81))
P5 = x(end,64) + 1600.*(x(end,6) + x(end,80))
Free_P5 = x(end,64)
P6 = x(end,69)
P7 = x(end,65)
P8 = x(end,67)
P9 = x(end,66)
P10 = x(end,63) + x(end,73)
P11 = x(end,71) + 14.*(x(end,78) + x(end,79) + x(end,80) + x(end,81))
Phage = x(end,74)
```

Load_Parameters.m

```
function [Parameters] = Load_Parameters(Master_Variables)
```

```
%This function will recalculate all parameters from a set of Master
%variables. The Master_Variables will be an array consisting of the
%following values:
```

```
%Master_Variables = [DNA Polymerase Elongation Rate(Nuc/s)      1
%                      RNA Polymerase Binding                    2
%                      RNA Polymerase Elongation Rate(Nuc/s)     3
%                      Ribosome Binding                          4
%                      Ribosome Elongation Rate(Nuc/s)           5
%                      mRNA A Degradation Half Life (minutes)    6
%                      mRNA B Degradation Half Life (minutes)    7
%                      mRNA C Degradation Half Life (minutes)    8
%                      mRNA D Degradation Half Life (minutes)    9
%                      mRNA E Degradation Half Life (minutes)   10
%                      mRNA F Degradation Half Life (minutes)   11
%                      mRNA G Degradation Half Life (minutes)   12
%                      mRNA H Degradation Half Life (minutes)   13
%                      mRNA Z Degradation Half Life (minutes)   14
%                      mRNA Y Degradation Half Life (minutes)   15
%                      mRNA W Degradation Half Life (minutes)   16
```

```

%-----Constants that Can Changes-----
%
%-----Replication-----
Length_Of_Genome = 6400;          %Nucleotides
%
%-----Transcription-----
PBSA = 0.6;          %Promoter Binding Strength of mRNA A Range 0.6-0.9
PBSB = 1.00;         %Promoter Binding Strength of mRNA B Range 1.00
PBSH = 0.65;         %Promoter Binding Strength of mRNA H Range 0.5-0.8
PBSZ = 0.02;         %Promoter Binding Strength of mRNA Z Range 0.15 - 0.30
PSBW = 0.225;        %Promoter Binding Strength of mRNA W Range 0.15 - 0.30
RNA_Length = 50;     %Number of base pairs RNA polymerase needs to travel to
                    %clear the promoter site
LA = 1900;           %Nucleotide Length of Transcript A
LB = 1000;           %Nucleotide Length of Transcript B
LH = 320;            %Nucleotide Length of Transcript H
LZ = 1600;           %Nucleotide Length of Transcript Z
LW = 1270;           %Nucleotide Length of Transcript W
%
%-----Translation-----
RBSP2 = 0.54;        %Ribosome Binding Site Strength of P2
RBSP10 = 0.054;      %Ribosome Binding Site Strength of P10
RBSP5 = 1.00;        %Ribosome Binding Site Strength of P5
RBSP9 = 0.08;        %Ribosome Binding Site Strength of P9
RBSP8 = 0.22;        %Ribosome Binding Site Strength of P8
RBSP3 = 0.40;        %Ribosome Binding Site Strength of P3
RBSP6 = 0.04;        %Ribosome Binding Site Strength of P6
RBSP1 = 0.08;        %Ribosome Binding Site Strength of P1
RBSP11 = 0.20;       %Ribosome Binding Site Strength of P11
RBSP4 = 0.23;        %Ribosome Binding Site Strength of P4
RSR = 173;           %Ribosome Spacing Requirement (Nucleotides)
LOP2 = 1200;         %Nucleotide Length of P2 Coding Region
LOP10 = 350;         %Nucleotide Length of P10 Coding Region
LOP5 = 270;          %Nucleotide Length of P5 Coding Region
LOP8 = 220;          %Nucleotide Length of P8 Coding Region
LOP3 = 1300;         %Nucleotide Length of P3 Coding Region
LOP6 = 340;          %Nucleotide Length of P6 Coding Region
LOP1 = 1040;         %Nucleotide Length of P1 Coding Region
LOP11 = 310;         %Nucleotide Length of P11 Coding Region
LOP4 = 1270;         %Nucleotide Length of P4 Coding Region
%
%
%
%-----Calculations-----
k_Replication = Master_Variables(1)/Length_Of_Genome;    %1/s
Ka = Master_Variables(2)*PBSA;                            %1/(s*molecule)
Kb = Master_Variables(2)*PBSB;                            %1/(s*molecule)
Kh = Master_Variables(2)*PBSH;                            %1/(s*molecule)

```

```

Kz = Master_Variables(2)*PBSZ; %1/(s*molecule)
Kw = Master_Variables(2)*PBSW; %1/(s*molecule)
RNAP_Sat = Master_Variables(3)/RNA_Length; %1/s
RNAP_A = Master_Variables(3)/(LA-RNA_Length); %1/s
RNAP_B = Master_Variables(3)/(LB-RNA_Length); %1/s
RNAP_H = Master_Variables(3)/(LH-RNA_Length); %1/s
RNAP_Z = Master_Variables(3)/(LZ-RNA_Length); %1/s
RNAP_W = Master_Variables(3)/(LW-RNA_Length); %1/s
KP2 = Master_Variables(4)*RBSP2; %1/(s*molecule)
KP10 = Master_Variables(4)*RBSP10; %1/(s*molecule)
KP5 = Master_Variables(4)*RBSP5; %1/(s*molecule)
KP9 = Master_Variables(4)*RBSP9; %1/(s*molecule)
KP8 = Master_Variables(4)*RBSP8; %1/(s*molecule)
KP3 = Master_Variables(4)*RBSP3; %1/(s*molecule)
KP6 = Master_Variables(4)*RBSP6; %1/(s*molecule)
KP1 = Master_Variables(4)*RBSP1; %1/(s*molecule)
KP11 = Master_Variables(4)*RBSP11; %1/(s*molecule)
KP4 = Master_Variables(4)*RBSP4; %1/(s*molecule)
Rib_Sat = Master_Variables(5)/RSR; %1/s
KEL2 = Master_Variables(5)/(LOP2-RSR); %1/s
KEL10 = Master_Variables(5)/(LOP10-RSR); %1/s
KEL5 = Master_Variables(5)/(LOP5-RSR); %1/s
KEL8 = Master_Variables(5)/(LOP8-RSR); %1/s
KEL3 = Master_Variables(5)/(LOP3-RSR); %1/s
KEL6 = Master_Variables(5)/(LOP6-RSR); %1/s
KEL1 = Master_Variables(5)/(LOP1-RSR); %1/s
KEL11 = Master_Variables(5)/(LOP11-RSR); %1/s
KEL4 = Master_Variables(5)/(LOP4-RSR); %1/s
Ads = log(2)/(Master_Variables(6)*60); %1/s
Bds = log(2)/(Master_Variables(7)*60); %1/s
Cds = log(2)/(Master_Variables(8)*60); %1/s
Dds = 0.7*log(2)/(Master_Variables(9)*60); %1/s
Dds_Alt = 0.3*log(2)/(Master_Variables(9)*60); %1/s
Eds = 0.7*log(2)/(Master_Variables(10)*60); %1/s
Eds_Alt = 0.3*log(2)/(Master_Variables(10)*60); %1/s
Fds = 0.7*log(2)/(Master_Variables(11)*60); %1/s
Fds_Alt = 0.3*log(2)/(Master_Variables(11)*60); %1/s
Gds = 0.7*log(2)/(Master_Variables(12)*60); %1/s
Gds_Alt = 0.3*log(2)/(Master_Variables(12)*60); %1/s
Hds = log(2)/(Master_Variables(13)*60); %1/s
Zds = log(2)/(Master_Variables(14)*60); %1/s
Yds = log(2)/(Master_Variables(15)*60); %1/s
YWs = log(2)/(Master_Variables(16)*60); %1/s

```

```

%-----
Parameters = [
k_Replication
Ka
Kb
Kh
Kz
Kw
RNAP_Sat
RNAP_A
RNAP_B
RNAP_H
RNAP_Z

```

```

RNAP_W
KP2
KP10
KP5
KP9
KP8
KP3
KP6
KP1
KP11
KP4
Rib_Sat
KEL2
KEL10
KEL5
KEL8
KEL3
KEL6
KEL1
KEL11
KEL4
Ads
Bds
Cds
Dds
Dds_Alt
Eds
Eds_Alt
Fds
Fds_Alt
Gds
Gds_Alt
Hds
Zds
Yds
YWs ];

```

Plot Phage Species

```

function [] = Plot_Phage_Species(WhichPlot,t,x)

%-----Documentation-----
%This file is used to plot the designed model species as a function of time
%The Input is a vector containing zeroes and ones. If it is equal to 1 then
%the group of species will be plotted. If zero than it will not be plotted.

%The Input is as follows:
%WhichPlot = [
%               All DNA species will be Plotted
%               Free RNAP promoter sites
%               RNAP promoter sites bound to a RNA polymerase
%               Elongating RNAP away from promoter site
%               Polycistronic mRNA species

```

```

%           Free Ribosome Binding sites
%           Ribosome binding site with ribosome attached
%           Elongating Ribosomes away from ribosome binding site
%           Phage Proteins
%           New Phage
%           Assembly Site formation species
%           E.Coli Proteins ]
%t = simulation time vector (In seconds)
%x = Matrix of phage species as a function of time
%-----
%-----

xaxis = ['Time [Seconds]'];

%Plotting All DNA Species
if WhichPlot(1) == 1
    Species_Name = {'ssDNA'; 'ssPDNA'; 'RF1'; 'RF2'; 'RF2DP3'; 'P5DNA'};
    for n = 1:6
        figure
        plot(t,x(:,n))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting all free RNAP promoter sites
if WhichPlot(2) == 1
    Species_Name = {'DA'; 'DB'; 'DH'; 'DZ'; 'DW'};
    for n = 1:5
        figure
        plot(t,x(:,n+6))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting all promoter sites occupied by an RNAP enzyme
if WhichPlot(3) == 1
    Species_Name = {'EA'; 'EB'; 'EH'; 'EZ'; 'EW'};
    for n = 1:5
        figure
        plot(t,x(:,n+11))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting all RNAP enzymes that are away from promoter site but still

```

```

%elongating on the mRNA
if WhichPlot(4) == 1
    Species_Name = {'ELA'; 'ELB'; 'ELH'; 'ELZ'; 'ELW'};
    for n = 1:5
        figure
        plot(t,x(:,n+16))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting All mRNA species
if WhichPlot(5) == 1
    Species_Name = {'A'; 'B'; 'C'; 'D'; 'E'; 'F'; 'G'; 'H'; 'Z'; 'Y'; 'W'};
    for n = 1:11
        figure
        plot(t,x(:,n+21))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting all free ribosome binding sites
if WhichPlot(6) == 1
    Species_Name = {'RBS2'; 'RBS10'; 'RBS5'; 'RBS9'; 'RBS8'; 'RBS3'; 'RBS6'; ...
        'RBS1'; 'RBS11'; 'RBS4'};
    for n = 1:10
        figure
        plot(t,x(:,n+32))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting all ribosome binding site with a ribosome attached to it
if WhichPlot(7) == 1
    Species_Name = {'RBS2R'; 'RBS10R'; 'RBS5R'; 'RBS9R'; 'RBS8R'; 'RBS3R'; ...
        'RBS6R'; 'RBS1R'; 'RBS11R'; 'RBS4R'};
    for n = 1:10
        figure
        plot(t,x(:,n+42))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting all ribosome still making protein and bound to mRNA
if WhichPlot(8) == 1

```



```

Species_Name = {'PD2'; 'PD10'; 'PD5'; 'PD8'; 'PD3'; 'PD6'; 'PD1'; 'PD11'; 'PD4'};
    for n = 1:9
        figure
        plot(t,x(:,n+52))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting ALL phage proteins
if WhichPlot(9) == 1
    Species_Name = {'P2'; 'P10'; 'P5'; 'P7'; 'P9'; 'P8'; 'P3'; 'P6'; 'P1'; 'P11'; ...
        'P4'; 'P2P10'};
    for n = 1:12
        figure
        plot(t,x(:,n+61))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting New Phage
if WhichPlot(10) == 1
    Species_Name = {'Phage'};
    for n = 1:1
        figure
        plot(t,x(:,n+73))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting Assembly Site Proteins
if WhichPlot(11) == 1
    Species_Name = {'Free Assembly Site'; 'Phage Initiation'; ...
        'Phage Elongation'; 'Phage Termination' };
    for n = 1:4
        figure
        plot(t,x(:,n+77))
        xlabel(xaxis)
        ylabel(Species_Name(n))
        grid

    end
end

%Plotting E.Coli Proteins
if WhichPlot(12) == 1
    Species_Name = {'DNA Polymerase'; 'Ribosomes'; 'RNA Polymerase'};

```

```
for n = 1:3
    figure
    plot(t,x(:,n+74))
    xlabel(xaxis)
    ylabel(Species_Name(n))
    grid

end

end
```

Appendix D: M13 Genome Sequence

The M13KE genome sequence was used for the input into the Salis Lab ribosome binding site calculator [28, 29]. The M13KE is identical to wild-type except it has an lac Z alpha gene located in the intergenic region. I used M13KE genome as the input to the ribosome binding site calculator because it was the one which immediately available to me and I do not anticipate to see a significant difference in using M13KE genome or wild-type. The whole genome, which consists of 7,222 nucleotides, could not be submitted for analysis at one time. Sections of the complete genome were submitted for analysis and prediction of the ribosome binding site strength. The only condition we held when creating when submitting the pieces of the genetic code is that at least 100 nucleotides were entered before the start codon.

LIST OF ABBREVIATIONS AND INITIAL PARAMETER VALUES

Symbol	Description	Initial Value*
ssDNA	viral single-stranded DNA	1
ssPDNA	ssDNA/DP3 binding complex	0
RF1	supercoiled double-stranded DNA with no breaks in (+) strand	0
RF2	relaxed double-stranded DNA with a break in the (+) strand	0
RF2DP3	RF2/DP3 binding complex	0
P5DNA	ssDNA sequestered by approximately 1600 P5 molecules	0
Di	free promoter site for the synthesis of mRNA i	0
Ei	RNAP bound to promoter site Di, inhibiting another RNAP from binding to it	0
ELi	RNAP transcribing mRNA i that does not occupy the promoter site	0
i	can be A, B, H, W, or Z	n/a
A	Polycistronic transcript coding for P2, P10, P5,P7, P9, P8	0
B	Polycistronic transcript coding for P10, P5,P7, P9, P8	0
C	Polycistronic transcript coding for P5,P7, P9, P8	0
D	Polycistronic transcript coding for P5,P7, P9, P8	0
E	Polycistronic transcript coding for P5,P7, P9, P8	0
F	Polycistronic transcript coding for P5,P7, P9, P8	0
G	Polycistronic transcript coding for P9, P8	0
H	Polycistronic transcript coding for P9, P8	0
Z	Polycistronic transcript coding for P3, P6	0
Y	Polycistronic transcript coding for P3, P6, P1, P11	0
W	Polycistronic transcript coding for P4	0
RBSj	Free ribosome binding site for translation of Pj	0
RBSjR	Ribosome bound to ribosome binding site j	0
PDj	Ribosome elongating protein j but not occupying the RBSj	0
j	can be any number between 1-11	n/a
P1	Assembly protein	0
P2	Control protein	0
P3	Coat protein	5
P4	Assembly protein	0
P5	Coat protein	0
P6	Coat protein	5
P7	Coat protein	5
P8	Coat protein	2700
P9	Coat protein	5
P10	Control protein	0
P11	Assembly protein	0
P2P10	P2/P10 binding complex	0
A _s	Free assembly site only containing the assembly proteins	0

P _I	Assembly site with 5 each of P7 and P9 bound to it	0
P _E	Assembly site with 5 each of P7 and P9, and 1 P5DNA bound to it	0
P _F	A newly synthesized phage particle still attached to the cell membrane	0
Phage	completed phage that have detached from the cell membrane	0
DP3	Host enzyme: DNA Polymerase III	3
RNAP	Host enzyme: RNA polymerase	1280
R	Host enzyme: Ribosomes	7880

*The initial value is the number of molecules of each species at the start of the infection (time = 0 seconds). The start of the infection corresponds to when the first ssDNA enters the cytoplasm.