2005

Engineering Genetic Networks to Measure Real Time Repression by λ Cro in vivo

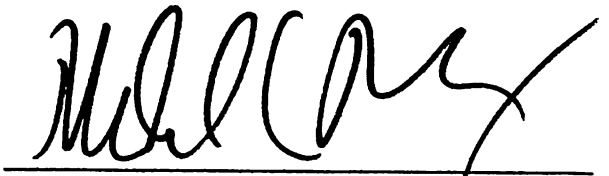
Zachary Louis LeBlanc

Engineering genetic networks to measure real time repression by λ Cro *in vivo*

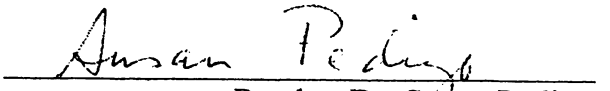
by
Zachary Louis LeBlanc

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell-Barksdale Honors College.

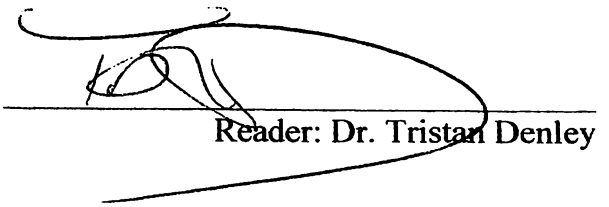
Oxford
May 2005



Advisor: Dr. Michael C. Mossing



Reader: Dr. Susan Pedigo



Reader: Dr. Tristan Denley

ACKNOWLEDGEMENTS

The author would like to acknowledge Dr. Michael C. Mossing, Associate Professor at the University of Mississippi, for his continued assistance and guidance throughout this project.

The author would like to acknowledge Yin Gu for her valuable assistance in cloning efforts.

The author would like to acknowledge Dr. Susan Pedigo and Dr. Tristan Denley for their participation in serving as readers of this thesis

ABSTRACT

ZACHARY LOUIS LEBLANC: Engineering Genetic Networks to Measure Real Time Repression by λ Cro *in vivo*

The field of systems biology seeks to engineer systems in order to explore specific dynamic properties as they act *in vivo*. Cellular activities have been proven to be regulated by a range of DNA binding proteins that bind to specific sequences that activate or repress transcription. Certain phage and bacterial repressor proteins were among the first of these DNA binding proteins to be characterized structurally and functionally. The demand for study of key steps in regulation of protein expression, folding, and repression has created multiple studies into systems of varying complexity. The data in this report explores the lambda system, and more specifically the kinetics of the repressor protein, Cro. The network of regulatory proteins and DNA sites that control a simple developmental switch in bacteriophage lambda have been extensively studied and modeled mathematically. Previous work in our laboratory has shown folding of Cro is very slow *in vitro*, and this phenomenon has not been included in models to date. The goal is to measure the time of Cro folding and repression *in vivo*.

In order to examine the kinetic properties of the Cro protein, a reporter plasmid containing the gene for the green fluorescent protein is fused to a promoter that is repressible by both Cro and lacR. Several strains of bacteria have been engineered containing lacR, the reporter plasmid, and optionally variants of Cro.

Initially GFP and Cro are expressed at a low level. After addition of IPTG, a compound that inactivates lacR, GFP accumulates and the fluorescence of the cell increases. In strains that contain the Cro gene, GFP accumulation is halted after enough active Cro has accumulated to repress transcription of GFP. Several Cro variants have been investigated. Initial observations of fluorescence spectra and induction kinetics of living bacterial cultures are reported.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I
Introduction: Molecular Biology 101

Watson and Crick's discovery of the structure of DNA (deoxyribonucleic acid) in 1953 unearthed the information pathway responsible for life at the molecular level. Prior to their groundbreaking discovery, Mendel first described the perpetuality of life as the passing of "traits" from progenitor to offspring through experiments with pea plants. Later, other scientists suspected DNA as the source of genetic information within all living cells. The work of Frederick Griffith and Oswald Avery proved that physical traits in bacterial cells could be transformed into cells that express a different phenotype, concluding that DNA must be responsible for the preservation and transfer of genetic information of every organism. The structure of DNA was the final piece of the puzzle, and has subsequently hatched a half century of scientific revolution (1).

DNA is composed of four different bases (adenine, thymine, cytosine, and guanine) grouped into two groups based on their ring structure (pyrimidines and purines). Each base is attached to a ribose sugar at the first carbon; a phosphate group is also attached at the fifth carbon on the sugar backbone. One unit of base, sugar, and phosphate is called a nucleotide. The sequence of DNA is formed from phosphodiester bonds between phosphate groups on one nucleotide and the hydroxyl group on the third carbon of ribose. These single stranded polymers form the double helix structure of DNA by bonding together through hydrogen bonds between

adjacent bases. The bases pair specifically to one another, purine to pyrimidine (adenine to thymine, guanine to cytosine) (2). Figure 1-1 shows the structure of DNA from building nucleotides to the double helix (2).

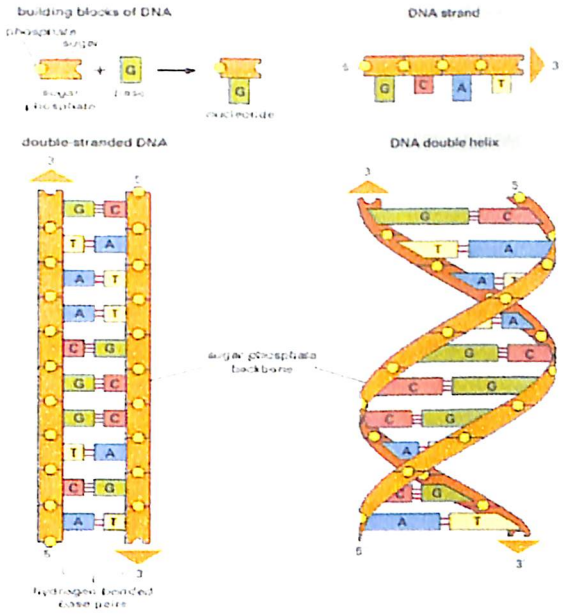


Figure 1-1: DNA structure. DNA is built from nucleotides units, containing ribose (sugar), a base (A,G, C, T), and a phosphate group. The single strand polymers wind together through hydrogen bonding between adjacent bases into the double helix structure. (Alberts *et al.* Molecular Biology of the Cell. Garland Science: New York, NY. 4th ed. 2002. Fig 4-3)

If DNA is molecule of information, then proteins convert that information into action. Proteins are responsible for several different activities inside living cells. First, they are the molecules that build membranes, organelles, and maintain the structure of DNA. Second, they can function as transport proteins, partially or completely embedded in cellular membranes. Lastly they act as enzymes, natural catalysts, which enhance the rate of cellular reactions. Protein structure focuses on amino acid units bonded together through peptide bonds. There are twenty different amino acids, each with specific chemical properties, which are critical to function of the folded protein (2).

The structure of proteins is best described in terms of primary, secondary, tertiary, and quaternary, as seen in Figure 1-2. Primary protein structure refers to the actual sequence of amino acids on the peptide chain. Secondary structure refers to

the folding, through chemical interactions within the sequence, into specific motifs. Tertiary structure is the final three dimensional structure of one specific polypeptide chain. The addition of other folded domains into a holistic, final structure is quaternary structure (2).

The code of DNA must be translated into a final protein product capable of doing work within the cell. The process of transcription describes the formulation of mRNA (messenger ribonucleic acid), a single stranded nucleic acid polymer similar to DNA. In prokaryotic cells, the RNA polymerase holoenzyme is responsible for unwinding the double helix and building a strand of mRNA using ribonucleotides. Ribonucleotides have the same sugar structure as DNA except for the presence of a hydroxyl group on the second carbon of the ribose. Additionally ribonucleotides do not have the base thymine, instead they substitute the base uracil. RNA polymerase gradually moves down a single strand of DNA and polymerizes the corresponding ribonucleotide to the mRNA strand (A for T, U for A, C for G, and G for C). Synthesis of mRNA begins at sections of the DNA celled promoters, which contain two specific sequences that allow for binding of RNA polymerase, the -10 and -35 boxes. RNA polymerase binds at these two sequences and begins transcription of mRNA at +1 on the strand of DNA and continues until a termination sequence in the DNA is reached causing release of RNA polymerase (2).

The completed mRNA transcript is then transported to special structure called ribosomes where translation occurs. Three sequential bases are called a codon and there are 64 known codons. Sixty-one of the 64 codons codes for a specific amino acid, the remaining three signal termination of translation. The ribosome reads the

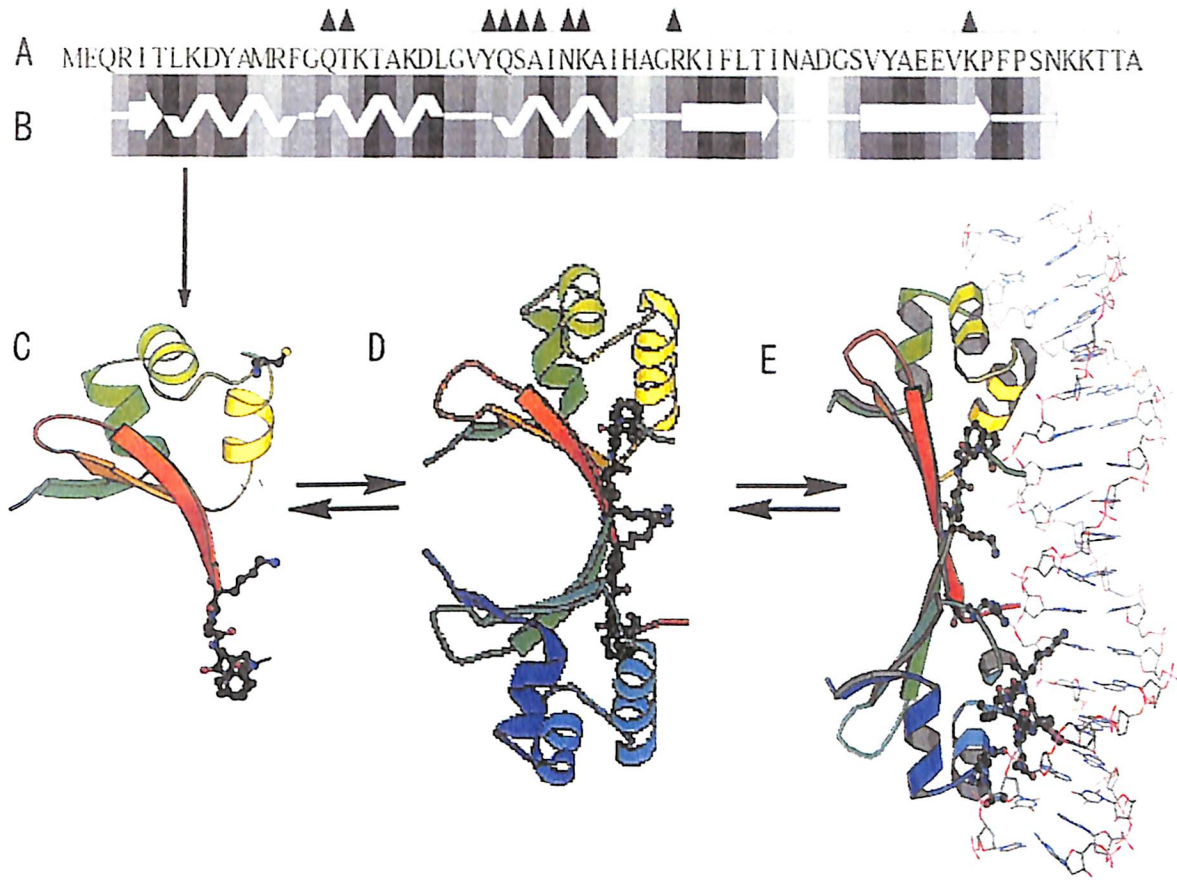


Figure 1-2: Cro structure. A is the amino acid sequence transcribed from DNA of Cro. The triangles indicate DNA binding points within the primary sequence. B is the secondary structure of Cro. White arrows indicate β pleated sheets and wavy lines indicted α helices. Tertiary structure is seen in C; this is representative of Cro monomer. D shows the dimmer bound together in quaternary structure. E shows the final step in the process of Cro, DNA binding. (Figure adapted from Albright *et al. Crystal structure of lambda-Cro bound to a consensus operator at 3.0 A resolution.* Journal of Molecular Biology: 280(1). pps 137-51.)

mRNA sequence in threes adding new amino acids and building a polypeptide chain. Release of the completed polypeptide chain results in folding and assembly of the chain into the final protein structure. These processes are often referred to as the central dogma of molecular biology and are represented in Figure 1-3.

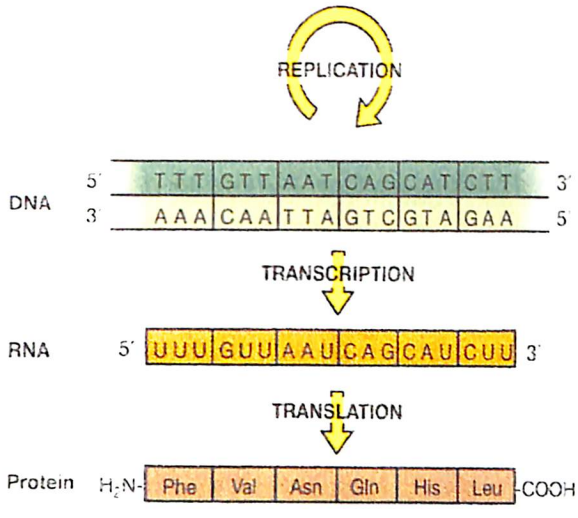


Figure 1-3: The Central Dogma – DNA is transcribed into mRNA (notice the substitution of U for T). The triplet codons, UUU for example, on the mRNA transcript each code for a specific amino acid. The mRNA transcript is translated into the primary structure of protein, a polypeptide chain (Phe, Val etc. are amino acids bound together through peptide bounds) (Madigan *et al.*. Brock Biology of Microorganisms. Prentice Hall: Upper Saddle River, NJ. 10th ed. 2003. Figure 7.1)

The process of transcription can be regulated by DNA binding proteins, folded to bind to specific sequences of DNA including the -10 and -35 regions of a promoter. Figure 1-2 shows the folded Cro protein bound to the double helix. This blocks binding of RNA polymerase and represses transcription. The specific sequences which bind these repressor proteins are called operators. The winding of the DNA double helix creates major and minor grooves. Watson and Crick's X-ray diffraction experiments deduced that there were approximately 10 base pairs per turn of the helix, with the major groove exposing more base pairs. DNA binding proteins are folded in such a way as to have specific domains that fit into the grooves of the double helix. The primary transcript of Cro from Figure 1-2 indicated specific amino acids that bound to DNA with triangles; these sites, when folded into the final protein, are exposed to specific nucleotides in these grooves. (2)

There are many different permutations of operators, promoters, and DNA binding proteins inside living organisms that regulate the behavior of that organism. There are examples of positive regulation such as the maltose regulon as shown in Figure 1-4. The maltose promoter demands an activator protein for polymerase binding. Maltose, a disaccharide composed of two glucose molecules, binds to this activator protein causing RNA polymerase to bind at the promoter. Then transcription of mRNA and translation of that information into folded proteins which digest maltose sugars in the cell. As long as there is maltose present in the cell, the proteins responsible for its digestion will be made because of the activator DNA binding protein and the orientation of the DNA sequence of the regulon. (4)

In other instances the DNA binding protein represses activity of RNA polymerase. An example of negative regulation is the lactose operon in Figure 1-4. This operon is composed of a promoter, an operator, and three downstream proteins that are responsible for lactose metabolism. Lactose is a disaccharide with a galactose and glucose molecule bound together. Bacterial cells produce lactose repressor that binds to the lactose operator; bound repressor prevents RNA polymerase from transcribing mRNA. Induction with lactose to the repressor releases it and allows transcription of the three genes. Therefore, the presence of lactose, like maltose, enables metabolism through regulation with DNA binding proteins. However, in this instance, the regulation is negative instead of positive. (4)

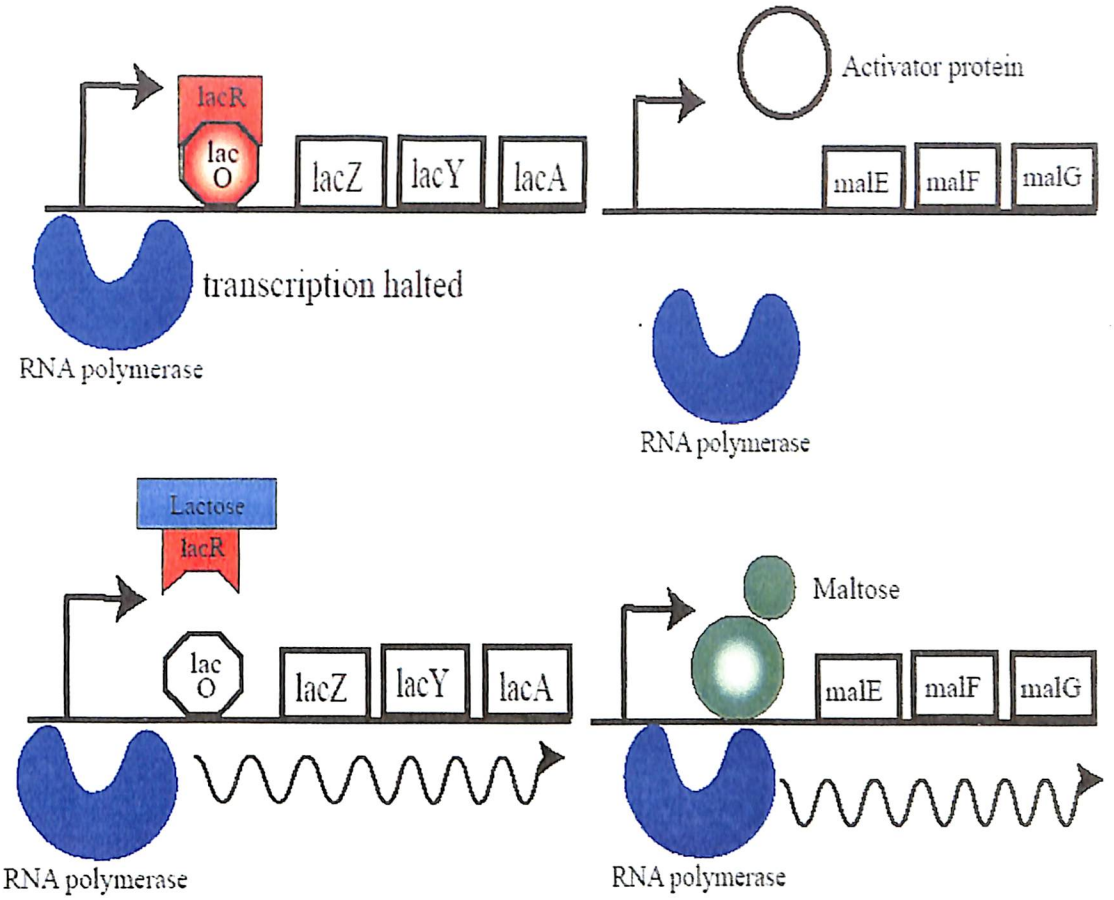


Figure 1-4: Regulation. The panels on the left are the lactose operon. LacR represses RNA polymerase through preventing binding at the promoter (bent arrow). Once lactose binds to lacR, repression is released and transcription proceeds. The downstream genes are responsible for lactose metabolism. The panels on the right are the maltose regulon. RNA polymerase can not bind without the presence of the activator protein. This protein can not bind to RNA polymerase without maltose. The need for maltose digestion causes binding of RNA polymerase to the promoter and transcription is the result. Again the downstream genes are responsible for maltose digestion. (Figure adapted from Madigan *et al.*. Brock Biology of Microorganisms. Prentice Hall: Upper Saddle River, NJ. 10th ed. 2003. Figs. 8.13 and 8.14)

The knowledge gained about the processes has lead to experiments to qualitatively and quantitatively measure these and other systems of regulation. Many questions about the kinetics of transcription, translation, protein folding, and DNA binding have been answered to date with numerous techniques *in vitro* or *in vivo*. Through these discoveries our scope of understanding becomes more and more specific, bringing us ever closer to understanding the fundamentals of life.

CHAPTER II
Systems Biology

Protein transcription, translation, and folding kinetics are at the heart of most cellular regulatory processes. In order for cells to function properly they must produce proteins to do the jobs necessary for life. Therefore it is not surprising the time necessary to produce fully functional proteins from DNA within cells has been extensively studied in classical *in vitro* experiments and lately *in vivo* with the aid of reporter genes and imaginative, often complex, engineered biological circuits. Reporter genes result in proteins, which are easily assayed and controlled. One product of the lactose operon, β-galactosidase, is a commonly used reporter because of its early establishment as a simple and reliable reporter as well as the convenience and sensitivity of its enzyme assay. However, in the past ten years green fluorescent protein (GFP) has become more popular in systems biology studies. GFP was isolated from the jellyfish *Aequorea victoria* by Shimomura *et al* in 1962 (5). The same group then measured fluorescence of GFP(5) along with many other discoveries in *A. Victoria* cells. However, it was not until the work of Chalfie *et al* that the full utility of GFP in molecular biology was realized; this group demonstrated GFP was fluorescent in prokaryotic and eukaryotic cells (5). Additionally GFP was proven to fluoresce in the absence of any additional jellyfish proteins (5). The green color is a result of a chemical reaction at the chromophore, a *p*-hydroxynenzlideneimid - azolinone formed at residues 65-67 in the folded protein (5). Brejc *et al* investigated

the folding of the normal chromophore present in wild type GFP compared to a serine substitution at position 65; in doing so, this group of scientists solved the structure of GFP through X-ray chrstallography (7). The structure of the folded protein, seen in Figure 2-1, is 11 β-sheets forming a hollow cylinder with a helix intertwined through the center. The chromophore, represented by ball and stick modeling in Figure 2-1, folds into the center of this helix (5). The neutral form of wild type GFP is excited at 470 nm and emits light at 504 nm, and folds efficiently, remaining stable at room temperature (6). Because jellyfish do not live at high temperatures, it has been shown that GFP is unstable at temperatures upwards of 65°C (6).

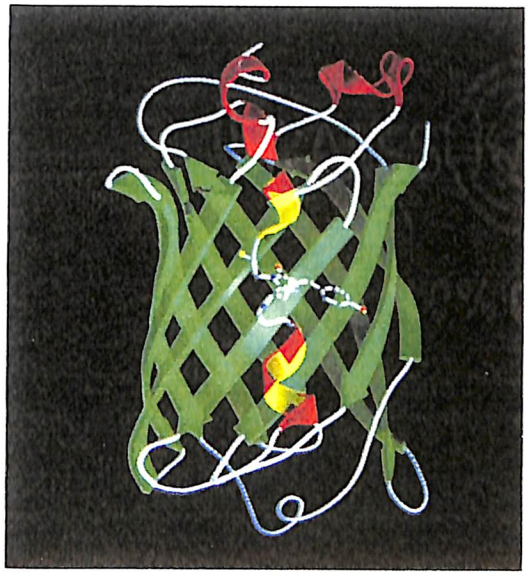


Figure 2-1: GFP structure. 11 β pleated sheets form a hollow cylinder where the chromophore (ball and stick modeling) is found. The figure is taken at 2.13 A resolution. (Brejc *et al. Structural basis for dual excitation and photoisomerization of the Aequorea Victoria green fluorescent protein*. PNAS: 94. p2306. 1997. Figure 1)

Relevant previous uses of GFP to this study include the work of Kalir *et al* on 14 flagella operons, measuring the temporal patterns of expression and further developing the model of flagella development (8). Setty *et al* demonstrated the use of GFP reporter plasmids in measuring the strength of the lac promoter in response to different combinations of cAMP and IPTG induction (9). The coupled use of mathematical modeling to map the lac promoter provided a level of detail *in vivo* not

yet seen in previous studies. By deriving a formula to determine the promoter strength under 100 different concentrations of IPTG and cAMP, they found the input function that determines the transcription rate of the operon (9). The definition of gene regulatory function (GRF), through study of the a synthetic λ cascade fused with florescent proteins as reporter genes, in Ronen *et al* resolved statistical noise in communication within cells. They studied the SOS system of DNA repair, isolating kinetic parameters of transcription from temporal measurements of promoter activity (10). The use of GFP in all of these studies allows for these types of measurements that are not possible in classical *in vitro* experiments. The current thought is the new field of bioengineering can answer questions of cellular regulation in any system in real time assays.

The λ "switch" is the most commonly studied regulatory mechanism to date because of the simplicity inherent in the study of viral particles compared to prokaryotes and eukaryotes. The switch describes the virus' decision between lysis and lysogeny, relating to which genes are turned on and which are repressed for the virus to survive. It is the nature of this decision that has made λ virus the paradigm for past, present, and future studies of cellular regulation. Therefore, much has been discovered about the elements of the switch, the roles played, and the kinetics involved from the in depth studies over the past 30 years. However, the advent of temporal measurements of protein kinetics, means new, more specific questions are surfacing.

This study specifically addresses the kinetics of the repressor protein, Cro; specifically the time needed for transcription, translation, folding, and DNA binding.

There are three DNA binding sequences (OR1, OR2, OR3) present in the λ operator, which controls the proteins CI and Cro. The presence of CI bound to OR1 and OR2 blocks transcription of Cro at the right promoter (P_R), preventing the production of new head and body proteins that are downstream of *cro*. The production of these proteins would lead to assembly of new viral particles and the eventual lysis of the cell. cI represses this activity and initiates the production of proteins that integrate λ genetic material into the bacterial chromosome, making it a bacteriophage. This life cycle is termed lysogeny. Cro binds to OR3 with the strongest affinity and repressors transcription of *cI* at P_{RM}. Cro also binds with less affinity to OR2 and OR1, repressing its own activity (11).

The P_R promoter of the λ genome is responsible for regulation of *cro* transcription. If RNA polymerase is allowed to bind to P_R within OR1 and OR2, the production of Cro will lead to the lysis decision. Previous studies of promoter strength have explored the best -10 and -35 sequences for Cro to bind. Takeda *et al* showed the use of base substitution experiments within P_R in order to determine the consensus sequence (13). This along with other in depth studies of the DNA binding domain of Cro have cultivated a knowledge base that makes this study not only possible, but the next logical step in the understanding of Cro kinetics.

The use of GFP fusions in study of the λ system has provided detailed analysis of the lysis-lysogeny decision in infected cells in Kobiler *et al* (14). GFP fusions to

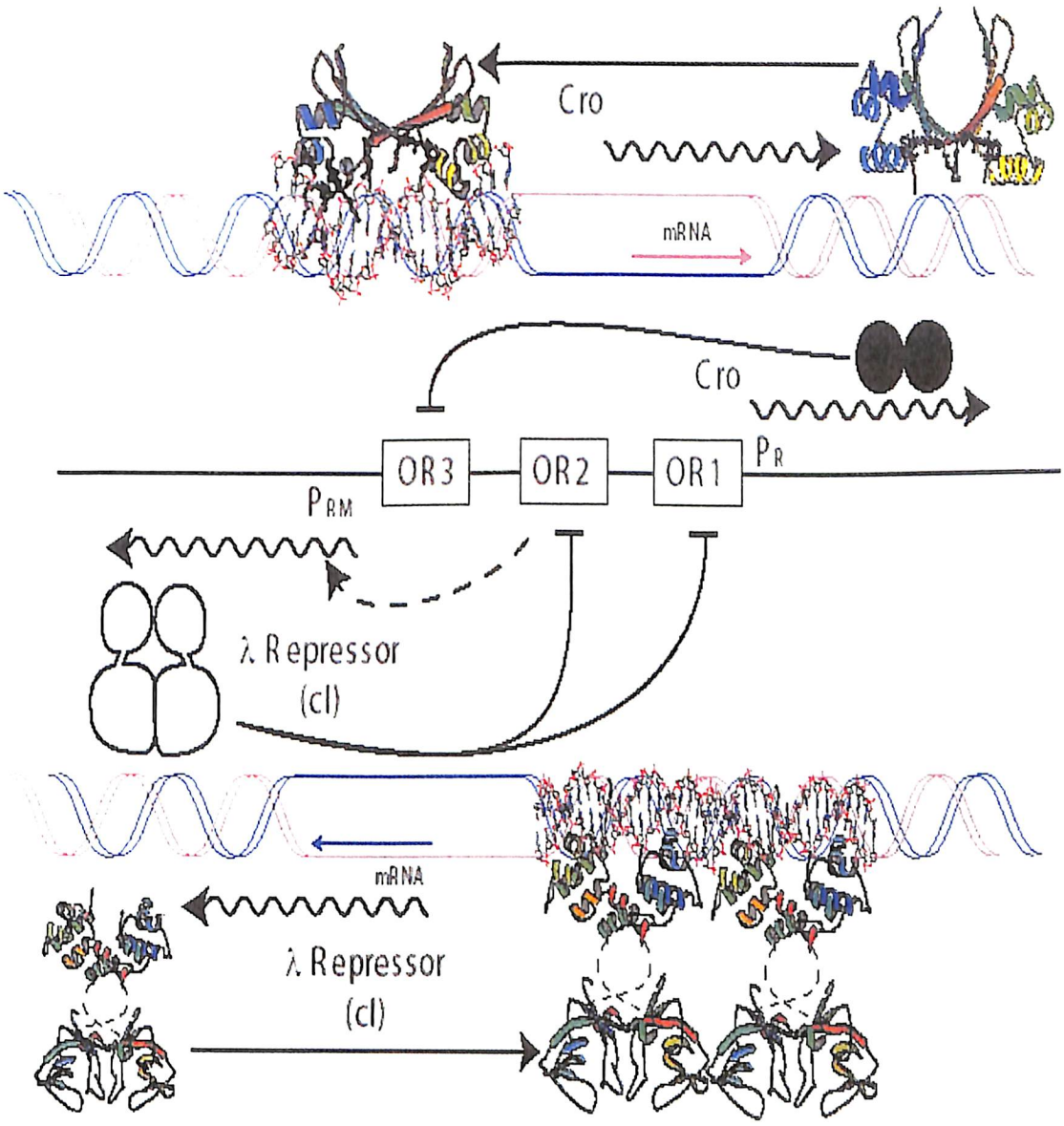


Figure 2-2: The λ switch. When Cro is present in the cell, it binds to OR3 to shut off expression of *cI*, and allowing the transcription of downstream genes responsible for the lysis life cycle. However, if CI is present, it binds OR1 to repress *cro* and OR2 to activate transcription at P_{RM}. This signals expression of genes downstream of *cI*, which are responsible for integration and the lysogeny life cycle. (Figure adapted Ptashne, Mark. A Genetic Switch: Phage lambda revisited. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. 3rd ed. 2004. Figure 3-1)

CII and Q protein were used to assay the finer details of *in vivo* regulation. Production of N protein is expressed and positively regulates the behavior of genes the left of P_{RM} and to the left of P_R. N activates *q* expression, which in turn activates expression of head and tail proteins late in the lysis life cylcle. CII production triggers RNA polymerase activity at P_{RE}, responsible for rightward expression of cI, and P_I, responsible for expression of *int* (11). When infected with only one phage particle, Cro was shown to reduce the activity of P_R and P_L promoters, inhibiting CII and promoting the lysis life cycle. However, in multi-phage infections, the expression of CII triggers expression of the repressor CI, inducing lysogeny and repressing Cro (14).

Figure 2-3: The λ map. Notice the DNA replication, head, and tail genes downstream of *cro* (clockwise). These genes are necessary for replication of the phage genome and packaging – prerequisites for lysis. The recombination genes, responsible for integration at the *att* site, are downstream of *cI* (counterclockwise). (Ptashne, Mark. A Genetic Switch: Phage lambda revisited. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. 3rd ed. 2004, Figure 3.1)

It is predicted that when different Cro plasmids are paired with a GFP reporter plasmid with a specialized form of P_R containing only OR1, that fluorescence assays will provide real time kinetic analysis of Cro folding and repression.

CHAPTER III
Methodology

The information contained in this chapter is a complete account of the genesis of this project, the early experiments in cloning, and the fluorescent measurements that were finally made. The groundwork laid in the past year has opened more questions than have been answered. It is the hope of the author that this chapter is of use to whomever continues this research.

The Control Sequence

From the beginning of this project the goal has been to develop an *in vivo* system that would measure the time from induction of Cro transcription to Cro repression. To accomplish this goal a control sequence that would regulate Cro and a reporter gene was necessary. The control sequence also had to be tightly repressed by Cro binding in order to ensure an accurate measurement. This control sequence also needed to express *cro* with the use of the λ operator sequences (OR1, OR2, OR3) and P_R from the λ switch. Additionally, the control sequence would express the reporter gene. Finally, lactose repressor would repress the entire system through its operator. The use of the lactose operon allows the system to be tightly repressed prior to induction, and easily induced with IPTG.

The λ operator sequences have very specific roles that have been proven experimentally. The molecular basis of each operator's affinity for both cI and Cro resides in the interaction between DNA and protein. For example, tight binding at

OR1 by cI blocks RNA polymerase binding at P_R and blocking transcription of *cro*. Additionally, cI acts to enhance its own activity through biding at OR2, which stimulates RNA polymerase to bind at P_{RM} to make more cI. However, unlike cI, Cro only represses expression negatively. Cro bound at OR3 represses activity at P_{RM}, the opposite effect of cI at OR2; Cro also represses its own activity through affinity with OR3. For survival of the virus when the cell is exposed to damaging events such at heat or UV light, it is important that flipping the switch first repress cI activity, which in turn allows for production of proteins crucial for viral reproduction and eventual lysis. For this reason, Cro binds most tightly to OR3 and least tightly at OR1. As the cell's machinery gradually makes viral proteins after the switch has been flipped, Cro buildup inside the cell eventually leads to complete occupation of OR3 and OR2, leaving OR1 open to shut expression of early lytic genes and *cro* (11). This gradual self repression of *cro* is seen in Figure 3-1.

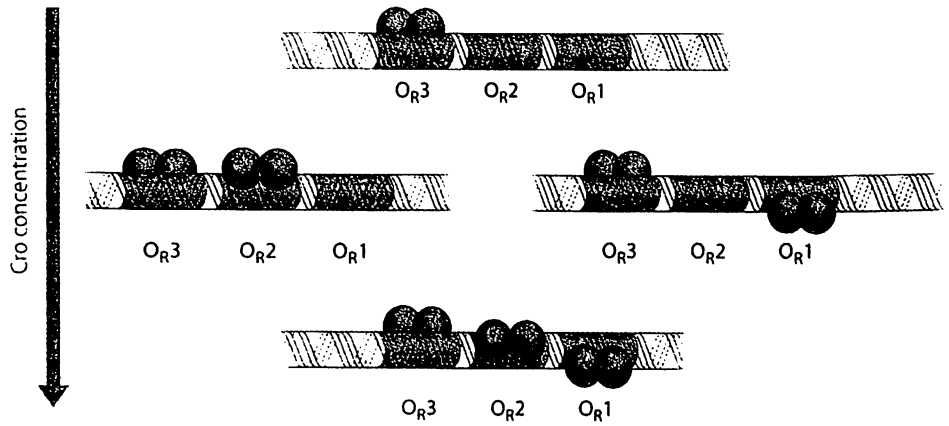


Figure 3-1: Operator saturation. As Cro dimers increase inside the cell, additional operator sites are occupied. Cro binds most strongly to OR3, repressing *cI*. More Cro dimers results in binding at OR2 and OR1, which blocks P_R and transcription of *cro*. (Ptashne, Mark. A Genetic Switch: Phage lambda revisited. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. 3rd ed. 2004, Figure 1.23)

With this understanding it was first decided that OR3 was ultimately irrelevant when examining *cro* repression. OR3 overlaps P_{RM}, responsible for expression of *cI* and not *cro*. Therefore, initial designs of the control sequence contained many combinations of OR2 and OR1 using the free energy measurements from Takeda *et al* (13). For simplicity sake, these plans were abandoned, opting for OR1 as the sole site of Cro binding. The OR1 sequence, flanked by the P_R -10 and -35 boxes, would allow accurate measurement of Cro repression *in vivo*. The addition of the lactose operator, which binds lactose repressor tightly, was the final piece of the control sequence. The complete control sequence is displayed in Figure 3-2. The +1 site of transcription initiation is present in the lactose operator (lacO).

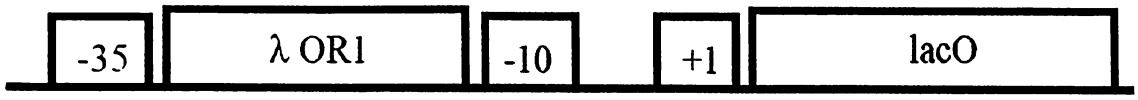


Figure 3-2: Pictorial display of the control sequence. -10 and -35 represent P_R, which surround OR1. Transcription begins at +1, just upstream of the lactose operator (lacO).

Viral Lysogeny

The first plan to engineer a system involved the creation of lysogen using phage infection. The control sequence would be attached to *lacZ* and *cro* and inserted into the λ genome. Furthermore, the genome would be engineered with the S12 (the streptomycin sensitivity gene). The DNA would then be packaged and used to infect a strain of lac$^-$ *E. coli*. Next, the viral DNA would integrate into the *E. coli* chromosome at the λ att site. Candidates would be chosen based upon blue growth on X-gal plates, assaying β-galactosidase activity, and growth in streptomycin media.

Cro dimmers repress the activity of S12, therefore integrates capable of Cro repression will grow in media (15). Induction with IPTG causes an increase of β-galactosidase activity until Cro was able to repress the activity or itself and β-galactosidase.

One advantage of this approach was the avoidance of multi copy plasmids which makes quantitative calculations of Cro imprecise. Additionally, the use of lysogens allowed for low levels of Cro in the cells, avoiding Cro toxicity. However, the complicated selection process, involving S12 and X-gal, provided uncontrollable variables and therefore allowed for too many variations in the final sequence in the *E. coli* chromosome.

mRNA Assay

If one could accurately measure the presence of *cro* mRNA inside a cell, the reporter gene could be eliminated. However, mRNA assays involve complicated laboratory technique with cDNA (16). Additionally, Cro repression, which occurs after transcription, would be only indirectly inferred without a real time assessment of Cro activity. Lastly, this plan was abandoned because of the inability of this system to make temporal measurements. Instead samples would have to be taken from induced cultures, translation halted with rifampicin, and then mRNA measured in each sample.

Plasmid recombination

A more straightforward method of making a measurement of Cro repression was to clone the control sequence in front of the reporter gene and *cro* on one plasmid. This plasmid would be transformed into competent *E. coli* cells that had the

lacI gene (responsible for expression of lactose repressor). The following methods were attempted.

Integration: The plasmid pAH125 from Professor Barry Wanner (Purdue University) is a β-galactosidase reporter plasmid capable of integration into the *E. coli* chromosome (17). First, the origin of replication on the plasmid, γ of R6K, is activated by *pir* to replicate. The *pir* gene was present on the chromosome of BW strains also acquired from Purdue University. Additionally, the λ attachment site (attλ) is on this plasmid, allowing for integration of the plasmid into the chromosome. Also a helper plasmid, pINT125, contained the integrase gene, which catalyzes the integration of pAH125 into the bacterial chromosome at attλ, completed the process as represented in Figure 3-4. pAH125 also had a polylinker region containing the restriction enzyme sites Sph1 and Xba1 as well as a kanamycin resistance gene; the map of pAH125 is in Figure 3-3.

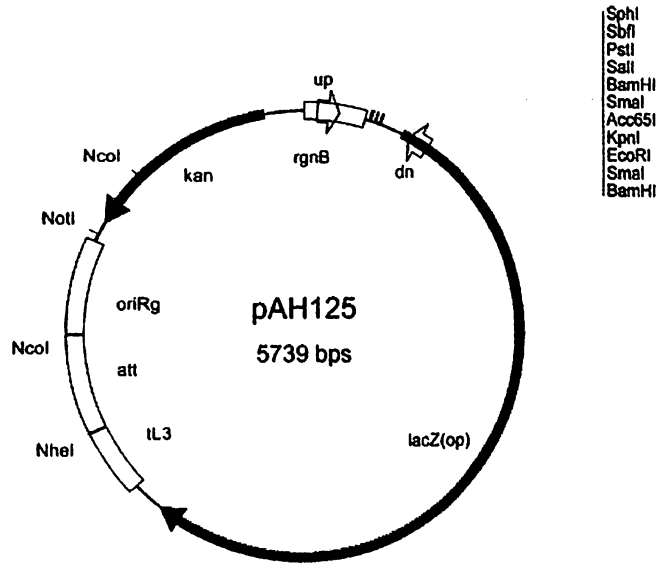


Figure 3-3: pAH125 map from Barry Wanner at Purdue University. The polylinker region is shown in the upper right corner. *lacZ* must have a promoter inserted for expression. (Haldiman and Wanner. *Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria.* Journal of Bacteriology: 183(21). pps. 6384-93. November 2001. Figure 5)

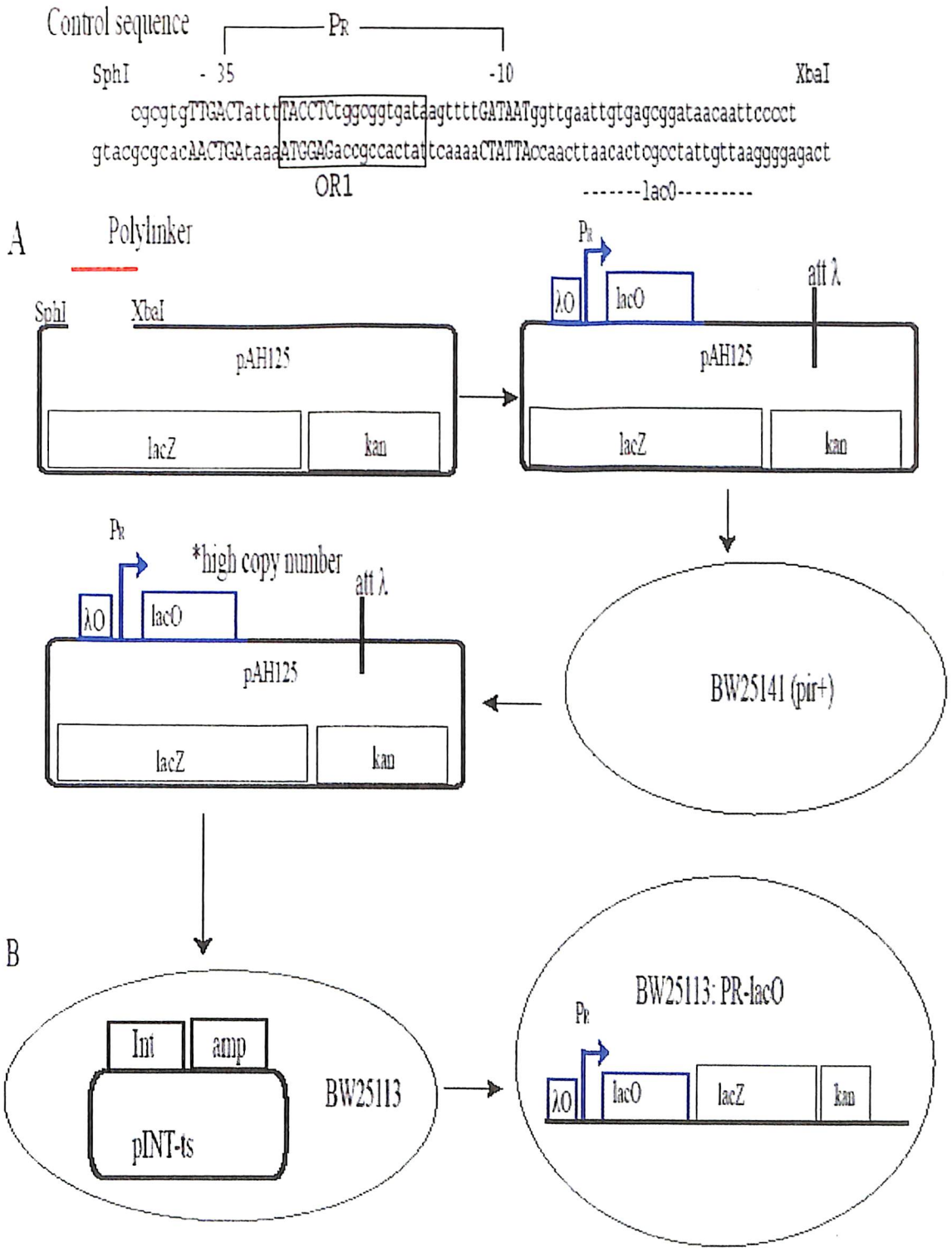


Figure 3-4: pAH125 integration. **A** represents excision of the polylinker sequence of pAH125 between restriction sites, SphI and XbaI. Next, P_R-lacO is ligated into pAH125 and the resulting plasmid is transformed into BW25141 (*pir*$^+$). Purification of pAH125 (P_R-lacO) from this strain gives a high yield of plasmid DNA. **B** represents transformation of pINT-ts and pAH125 (P_R-lacO) into BW25113 (*pir*$^-$). When the temperature is raised to 42°C activated the integrase on pINT-ts and pAH125 (P_R-lacO) intergrates onto the chromosome. **C** represents the transformation of pET-Cro variants into BW25113: P_R-lacO. The final system can now be induced with IPTG and β-galactosidase measurements taken.

The control sequence was ordered in two complete sequences (approximately 70 bp each) from Integrated DNA Technologies, Inc. with the restriction sites SphI and XbaI on either end of the oligonucleotide. Restriction sites are DNA sequences that can be cleaved by special restriction endonucelases. Their use makes plasmid cloning possible. The sequence was purified on MonoQ, dialyzed, annealed, and analyzed spectroscopicly ($\lambda = 595$ nm). Purification of P_R-lacO was necessary because of the instability of single stranded oligonucleotides over 60 base pairs long. Only approximately 60% of the ordered oligonucleotide is the full length. Purification eliminates smaller segments of oligonucleotides. The plasmid was then digested with Sph1 and Xba1 from New England Biolabs and analyzed by agarose gel electrophoresis. The gel showed a clear distinction between uncut plasmid and digested plasmid. The gel in Figure 3-5 shows digestion of pUA139 as an example of agarose gel electrophoresis. Next the annealed control sequence was ligated using T4 DNA ligase and transformed into JM109:DE3 strain of *E. coli*. Transformations were plated on LB kanamycin (50 μg/mL) plates in order to select for kanamycin resistance. However, no candidates for further study were observed.

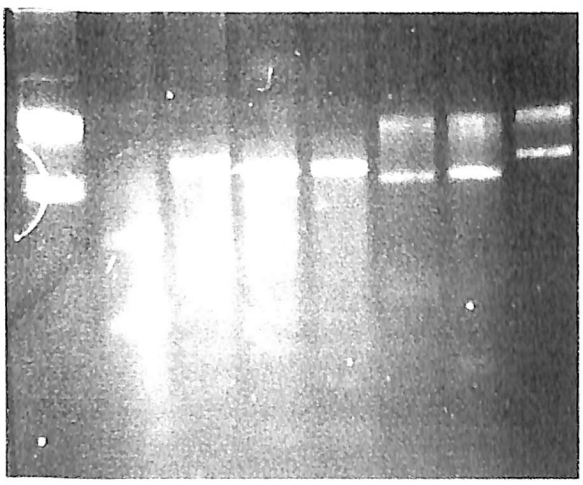


Figure 3-5: pUA139 digestion. Lane 1 (left) is undigested DNA. Lane 2 is digested with HaeIII. Lane 3 is digested with HincII. Lane 4 is digested with HincII and XhoI. Lane 5 is digested with XhoI. Lane 6 is digested with SphI. Lane 7 is digested with SphI and BamHI. Lane 8 is digested with BamHI. This gel shows complete digestion of pUA139 with HaeIII and HincII/XhoI. Digestion with SphI/BamHI is incomplete, as evidenced by a closed circular band aligned with the uncut DNA in Lanes 6, 7, and 8.

pET Cro plasmids: The pET vector supplied by New England Biolabs had been used in the lab previously to create different Cro variants. These variants included wild type Cro, single chain Cro (contains a linker connecting two *cro* genes, which makes folding much quicker), N-His Cro (normal *cro* gene tagged at the N-terminus with histidine), single chain single and double mutants (proline mutations on one or both *cro* genes). Each of these cloning vectors also had XbaI and SphI restriction sites as well as an XhoI site after the *cro* gene. The plan was to ligate the control sequence into each of these variants, creating P_R-lacO-cro (**A** from Figure 3-6). Then, digest with SphI and XhoI both pAH125 and the cloned pET-Cro vectors. The P_R-lacO-cro sequence would be purified from the remainder of the pET-Cro plasmid. The cut plasmid would be run on an agarose gel with an UV dye. The smaller fragment could be excised from the gel using a UV preparative and purified from the gel slice. The purified plasmid DNA would be ligated into pAH125 in front of *lacZ* (**B** from Figure 3-6). The resulting pAH125 (P_R-lacO-cro-lacZ) plasmid would be transformed into BW25131 (pir^+) to allow for high yield in plasmid purification (**C** from Figure 3-6). Purified plasmid would be transformed into the strain BW25114, already having pINT-ts. After raising the temperature, activating integrase on pINT-ts, the P_R-lacO-cro-lacZ segment would integrate onto the chromosome. The completed system would be induced with IPTG and β-galactosidase assayed (**D** from Figure 3-6). However, ligation of the control sequence into pET plasmids proved difficult and the P_R-lacO-cro sequence was never generated and transformation never attempted.

GFP: The pUA139 plasmid from Uri Alon has the GFP gene as its reporter and a kanamycin resistance gene (9). New oligonucleotides were ordered in four segments

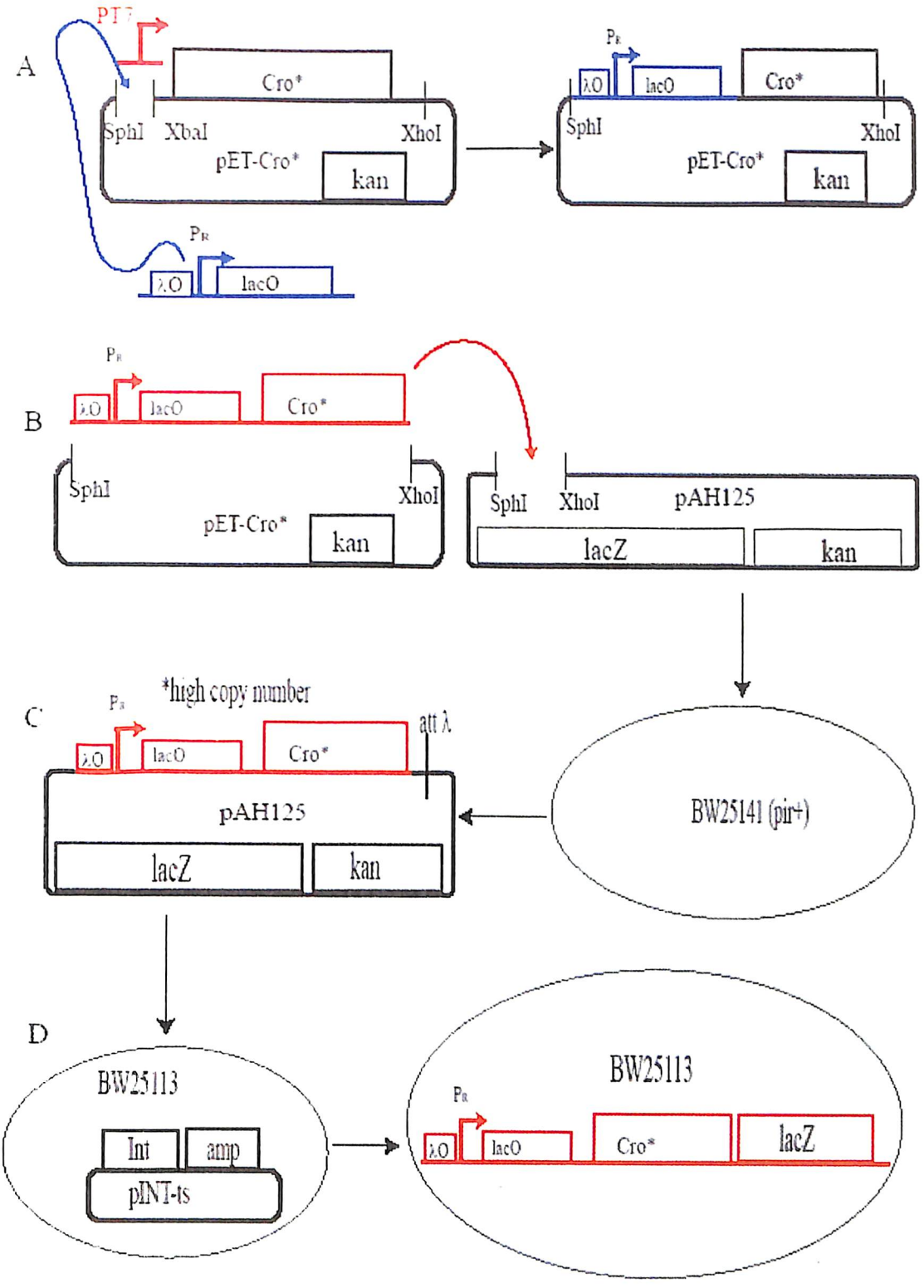


Figure 3-6: pETCro intergration. Integration of P_R-lacO-cro-lacZ into BW25113. The control sequence is cloned into pEtCro plasmids in between SphI and XbaI, replacing the T7 promoter in **A**. The control sequence fused to *cro*, between sites SphI and XhoI is removed and ligated into pAH125 in front of *lacZ* in **B**. **C** describes the purification of pAH125 (P_R-lacO-Cro) from BW25141 to achieve high yield of plasmid DNA. pAH125 (P_R-lacO-Cro) is then transformed into a *pir*$^+$ strain, BW25113 for integration into the chromosome with the helper plasmid, pINT-ts in **D**.

(T1, T2, B1, and B2) from IDT. These segments were kinased with polynucleotide kinase (PNK) and annealed. In addition, two new restriction sites were chosen and sequenced onto the ends of P_R-lacO – BamH1 and Xho1. Figure 3-7 is the sequence ordered for cloning from IDT. Both of these restriction sites were present in the polylinker of pUA139. pUA139 was then digested with BamH1 and Xho1 and analyzed with agarose gel electrophoresis to ensure complete digestion. Next, the kinased, annealed control sequence was ligated into pUA139 (**B** from Figure 3-8). The resulting plasmid, P_R-lacO-GFP, was then transformed into BW25113 and analyzed with fluorescent spectra (excitation of 485 nm and emission of 535 nm) on a Habira FluroMax-3 fluorometer (Figure 4-1).

```
   XhoI        [ -35]    lambda PR      [ -10]
5' gatccgcgtgTTGACTatttTACCTCtggcggtGATAATggttgaattgtgagcggataacaattcccc 3'
3'     gcgcacaactgataaaATGGAGaccgcCACTATtaccaacTTAACACTCGCCTATTGTTAAggggagct5'
                                                 --------lacO-------    BamHI
```

Figure 3-7: DNA sequence of PR-lacO-λO. Ordered from IDT.

A different set of Cro variants were chosen to accompany P_R-lacO-GFP in the real time experiments, pUCro plasmids. These plasmids were ampicllin resistant, which made isolation of colonies with both P_R-lacO-GFP (kan^R) and pUCro (amp^R) after transformation easy. The pUCro variants included wild type cro, K56 [DGEVK] (Cro monomer, incapable of binding DNA), F58W, and Y26C. These plasmids were present in X90 strain of *E. coli* at -70°C. Single colonies were streaked out from freezer stock onto LB ampicillin (100 μg/mL) plates and grown overnight at 37°C. Fresh single colonies were then resuspended in 100 μL 1X TSS buffer (5 g poly ethylene glycol, 2.5 mL 1M Mg^{2+}, fill to 50 mL with LB and filter

sterilize .2 μm). 10 ng of P_R-lacO-GFP was then added to the culture suspension and incubated on ice for 30 minutes. At which point, .9 mL 1X TSS was added. Then the culture were plated in a 42°C water bath for 1.5 minutes, incubated on ice for 5 minutes, and shaken at 200 rpm at 37°C. After an hour, the cells were plated on LB kanamycin (50 μg/mL) and ampicllin (100 μg/mL). After growth at 37°C overnight, transformants were picked, replated on LB – kan/amp plates and resuspended in M9 selective media (9) to grow at 37°C overnight. Complete cloning scheme represented in Figure 3-8.

Overnight cultures were diluted with M9 selective media to an optical density of .05 in a Falcon 3047, 24-well plate. Samples included wild type cro, Y26C, F58W, K56 [DGEVK], and P_R-lacO-GFP. Readings were taken in a Wallac 1420 plate reader, which took readings at room temperature. Each run measured absorbance at 570 nm and fluorescence (excitation 485 nm and emission 535 nm) of each well in the plate. After the measurements were taken, the plate was shaken for 30 seconds, and allowed to rest for 1 minute at which point another measurement was taken. The process would continue for two hours to allow the cells time to acclimate room temperature. At this point each, 100 mM IPTG was added to each well and another series of measurements were started. The data was exported to Matlab and the fluorescent and absorbance spectra were plotted.

Figure 3-9 describes the effect of induction upon the engineered system. Prior to induction lacR is bound to lacO, preventing transcription of both GFP and Cro. IPTG releases lacR allowing RNA polymerase to bind to P_R and P_{tac}. RNA polyermase can then make GFP and Cro mRNA, which will be folded into the final

protein structure. Folded GFP will fluoresce until enough Cro accumulates in the cell to bind OR1 and prevent further GFP expression. At this point the fluorescent readings should deviate from the unrepressed control.

Computer Simulation

In order to predict the behavior of the system after induction with IPTG, known kinetic parameters were used in the Dizzy program. This program developed at the Institute for Systems Biology for stochastically modeling genetic regulatory networks. Through modeling the kinetic behavior of the system, the real time data can be analyzed more precisely because of the reference provided by Dizzy (18).

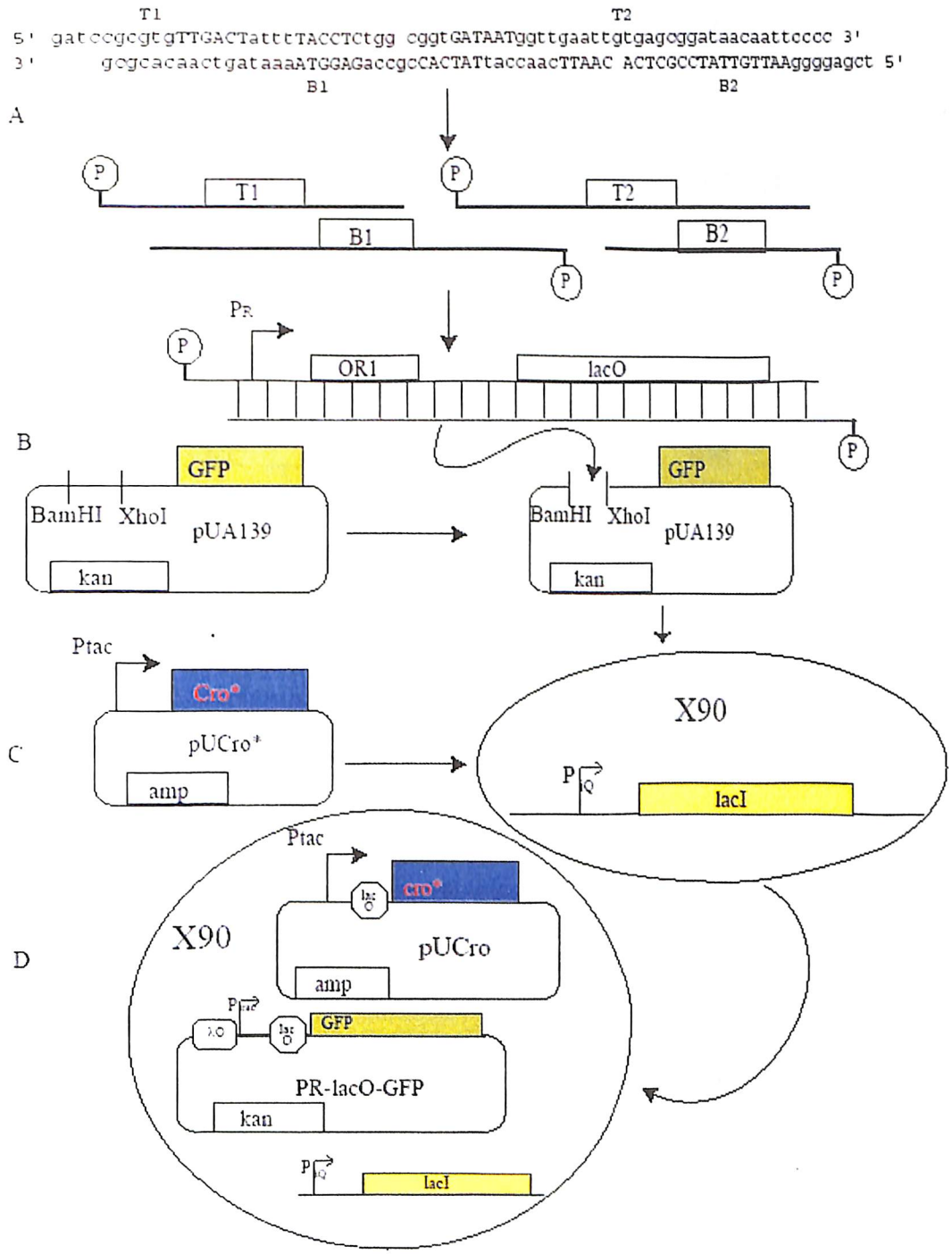


Figure 3-8: GFP cloning scheme. **A** - The control sequence was kinased with PNK and annealed. **B** – P_R-lacO was ligated into pUA139 digestion with XhoI and BamHI. **C** – Both pUA139 (P_R-lacO) and pUCro plasmid were transformed into X90 strain of *E. coli*. **D** – The final system with lacI (expression of lacR) on the chromosome, the reporter plasmid, and the Cro plasmid.

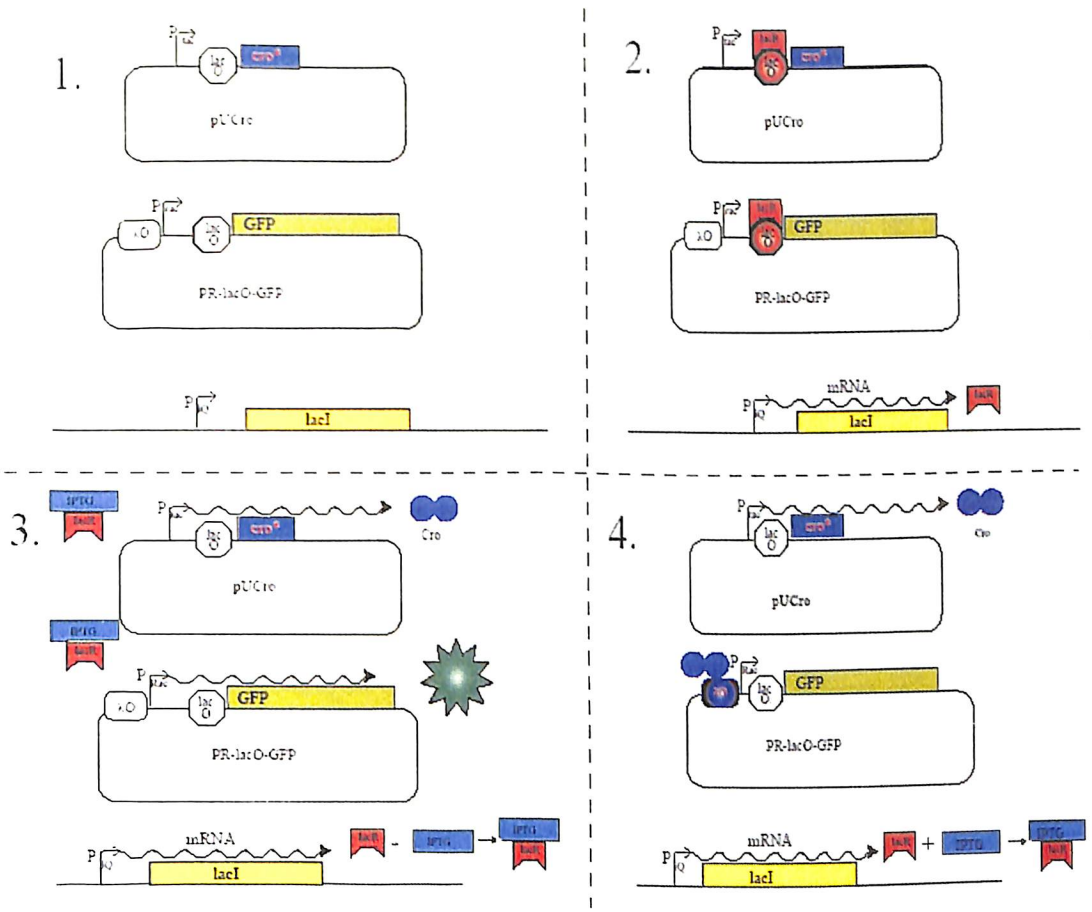


Figure 3-9: Final system. Shown above is process by which GFP fluorescence was assayed through repression by lacR (1), induction with IPTG (2), transcription of Cro and GFP (3), and final Cro repression (4).

CHAPTER IV
GFP Repression

Figure 4-1 shows the fluorescent spectra of pUA139 without P_R-lacO and with P_R-lacO-GFP uninduced and induced. The promoterless plasmid produced little fluorescence signal over the range detected. There is fluorescent characteristic of GFP in the uninduced case. However, the peak at 505 nm is three times less than the peak seen in the induced case. It is still noteworthy to indicate the background fluorescent signal detected in the cells without GFP production. Clearly all repressors are not perfect and lacR might only repress GFP activity 90%, which would result in a high background fluorescent level in the uninduced case. Also, GFP folding, assembly, and fluorescence might lag behind the actual real time data to some extent. This lag time could also explain a high background level fluorescence. To correct for this problem the P_R-lacO-GFP and pUCro plasmids were transformed into X90 strain of *E. coli* in the real time assays.

Figure 4-2 shows computer-simulated production of GFP under three different conditions. The first condition, without a Cro plasmid in the cell, shows an uninterrupted steadily increasing GFP profile. The GFP fluorescence from the spectra in 4-1 predicts this activity. The absence of Cro production to repress GFP results in normal fluorescence as in the peak from Figure 4-1.

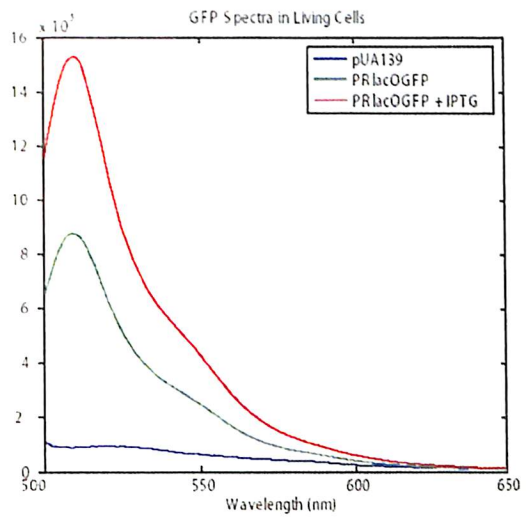


Figure 4-1: GFP Fluorescent spectra. Shown are pUA139, P_R-lacO-GFP, and P_R-lacO-GFP with IPTG. Measurement taken on J-Y Habira Spex Flouromax-3, excitation 485 nm and emission 500 – 650 nm. (Ramsey *et al. Dizzy: Stochastic Simulation of Large-Scale Genetic Regulatory Networks*. Journal of Bioinformatics and Computational Biology: 3(2). pps 415-36. 3 April 2005.)

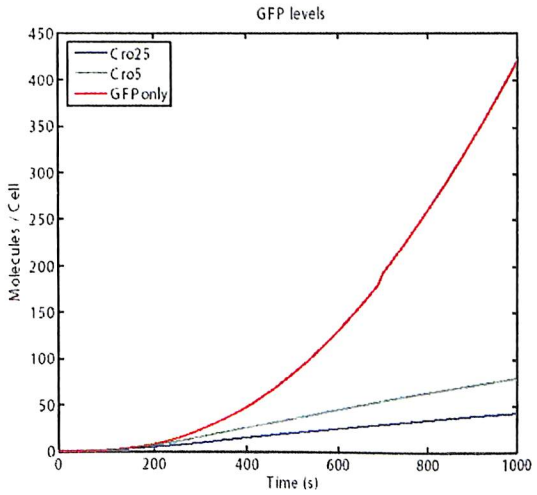


Figure 4-2: GFP simulation. Simulation showing GFP molecules per cell as a function of time. Cro 5 and Cro 25 refer to the copy number of the Cro plasmid in the cell. (Ramsey *et al. Dizzy: Stochastic Simulation of Large-Scale Genetic Regulatory Networks*. Journal of Bioinformatics and Computational Biology: 3(2). pps 415-36. 3 April 2005.)

The second condition, Cro5 in Figure 4-2, represents a Cro plasmid at a copy number of five. In this condition there is a decreased GFP profile compared with the uninhibited simulation. The third condition, with a Cro plasmid of copy number twenty-five, results in GFP levels less than the simulation with five Cro plasmids. Copy number refers to the amount of times the plasmid will replicate itself before the bacterial cell divides. The trend here, increasing Cro decreases the amount of GFP in the cell significantly, is expected. As Cro builds up in the cell, especially at a 25:1 ratio to GFP, GFP expression is repressed at the λO site on the control sequence and fluorescence is predicted to deviate from the condition without Cro.

Figure 4-3 shows the production of Cro and Cro dimer in relation to GFP levels in the cell at a copy number of twenty five. The production of Cro in this simulation is seventeen times higher than the production of GFP, and molecules of Cro dimer is ten times more. This confirms the assertion from Figure 4-2 that over production of Cro by high copy number pUCro plasmids causes a drastic decrease in GFP molecules in one cell.

Figure 4-4 shows the real time data from the Wallac 1420 plate reader. As seen in the computer simulation, in the absence of Cro, fluorescence due to GFP shows an uninhibited increase in fluorescence in Panel A. This increase shows no signs of repression because of a lack of Cro in the cells. Additionally, induction with IPTG shows no significant effect on the growth rate of the cells (Panel B) because the absorbance spectra show similar rates of growth, increasingly at a steady rate.

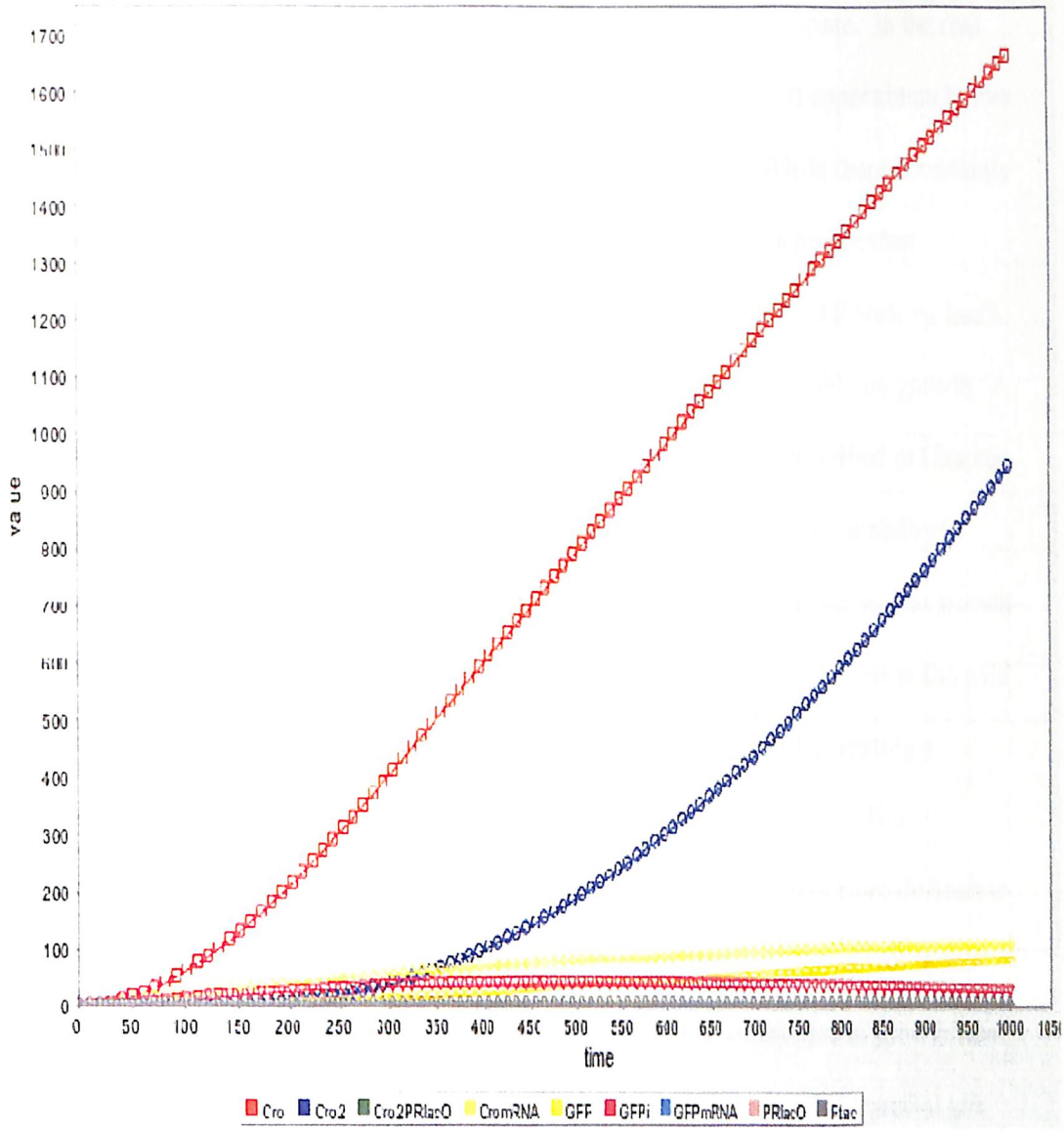


Figure 4-3: Cro simulation. Simulation showing Cro (copy number 25) production in the same simulation as Figure 4-2. The red line, Cro prior to dimerization, and the blue line, Cro dimmer, is far above the production of GFP molecules. The repression of GFP overshadows the expected fluorescent profile from the system. (Ramsey *et al. Dizzy: Stochastic Simulation of Large-Scale Genetic Regulatory Networks.* Journal of Bioinformatics and Computational Biology: 3(2). pps 415-36. 3 April 2005.)

Furthermore, the use of GFP as a reporter gene seems to be appropriate. In the real time assays the promotorless wells remained at a background level approximately two times less than GFP with or without pUCro (data not shown). While there is certainly some background fluorescence, whether from leaky GFP expression or other fluorescent molecules in the cell, the elevated fluorescent level of GFP with P_R-lacO-GFP remained constant. GFP also appears to have no significant effect on growth rate of the cells. Lastly, GFP is easily measured by the protocol described in Chapter III in the plate reader. The spectra in panel C indicate Cro monomer's inability to repress GFP fluorescence versus the uninduced state (Panel D). However, the growth rate of the induced cells is less than the uninduced. This trend is also seen in the wild type Cro, albeit more dramatic. Cro inhibits the growth of the cells, creating a different absorbance profile for the cells induced with IPTG (Panel F). The fluorescence timecourse (Panel E) for the wild type plasmids are therefore difficult to interpret. Cro, a protein that normally signals lysis of the cell and is subject to self-repression, has unrepressed expression in this system. The difference in growth rate can be directly attributed to the presence of Cro in the cells, as the GFP control saw no difference in growth rate as a result of induction. This system is designed so that, in a short period of time, Cro overwhelms the cells biochemical machinery, upsetting the delicate balance within the cell. Replication, transcription, and translation enzymes seem to become overwhelmed early after induction creating an unhealthy metabolic environment for the cells that inhibits growth. This was predicted in the computer models when the presence of Cro dimer is ten times more than GFP, this

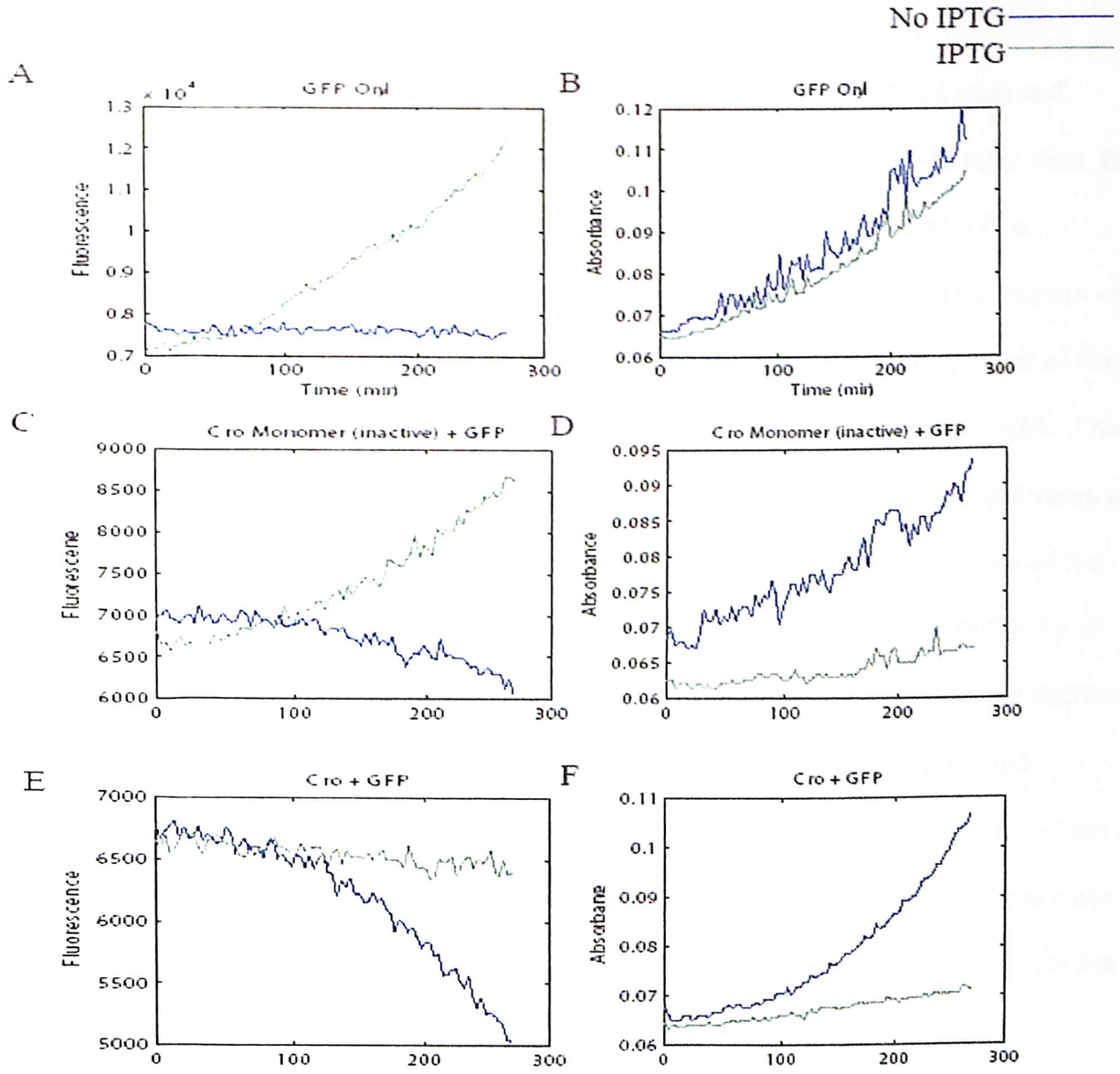


Figure 4-4: Real time data. Fluorescence (GFP) and Absorbance (cell density) of cell cultures as a function of time after induction. All plots indicate an average of values over time. Date collected in the Wallac 1420 plate reader and compiled in Matlab. **A** - fluorescence date for cells containing P_R-lacO-GFP. **B** – absorbance for cells containing P_R-lacO-GFP. **C** – fluorescence date for cells containing P_R-lacO-GFP and pUCro monomer. **D** - – absorbance for cells containing P_R-lacO-GFP and pUCro monomer. **E** - fluorescence date for cells containing P_R-lacO-GFP and pUCro **F** - absorbance for cells containing P_R-lacO-GFP and pUCro.

trend can reasonably be extrapolated to other proteins critical to cell vitality within the cell.

The significant decline in fluorescence signal in the uninduced state and consistent data of induced state are indicative of this abnormal growth data. First, the decrease in fluorescence of the induced case could be the result of leaky Cro expression. If, as predicted above, the background GFP levels seen in the spectra are due to incomplete binding of lacR to the operator sequence, then any number of Cro molecules that are allowed to fold and bind to λO will repress GFP expression. This is unlikely, because the cells seem to grow at a rate similar to the uninduced states of GFP control. However, the decrease in fluorescence is also seen in the case of the monomer without IPTG. It should also be noted that the decrease of 2000 units on the graph of wild type Cro is twice that of the monomer decrease. Also, the decrease begins at the same time, 100 minutes after induction, as the control GFP and monomer began their increase in fluorescence. Perhaps the decrease in fluorescent is evidence of dilution of GFP molecules in the cells. The pUA139 plasmid has a copy number of one, and the increasing number of GFP is not able to compensate for the increase in growth.

The other anomaly in the wild type Cro fluorescent panel is the induced state does not increase. Recall that the computer simulation showed a sudden increase, without the hypothesized lag in repression, with Cro in the cell. Analysis of Figure 4-3 shows that within 50 minutes Cro levels are already one hundred times that of GFP. The overwhelming number of unrepressed Cro molecules in the cells, not only affects

the growth rate of the cells, but also hides the real time expression of GFP. In comparison to the monomer, which does not bind λO, wild type Cro is already at a steady state of GFP repression.

Finally, there is a difference in the fluorescent spectra of uninduced and induced samples. Therefore, repression of GFP activity is seen by wild type Cro. In the other data samples of Cro variants the affects of repression result in slower increases of fluorescence compared with the GFP control. This data is hard to reproduce and therefore the actual difference between the mutants, wild type Cro, and control GFP hard to estimate. It can be said qualitatively that Cro does repress GFP, although quantitative analysis would be unreliable.

CHAPTER V
Plans for the future

This project has made a measurement of *in vivo* regulation using an engineered system. However, after analysis of the data the flaws in the system are obvious. Cro is allowed to be expressed in abundance without self repression which results in poor growth and unreliable GFP repression. Therefore, the system needs to be redesigned to compensate for the initial model's insufficiencies. It seems that through the fluorescent spectra and real time data that GFP is a relatively stable, quick folding, reliable reporter gene that is easy to assay with the Wallac 1420. Future work should pursue cloning efforts to place Cro, GFP, and the control sequence on one plasmid. Additionally new control sequences have been ordered with different λO sequences. These include a sequence that is an exact consensus sequence, allowing for tight Cro binding, and another with several point mutations that will not allow wild type Cro binding. Use of these operator sequences could offer more evidence of Cro repression *in vivo*. With the use of the current GFP plasmid, the current control sequence, and these unused operators questions posed by these results could certainly be resolved.

Another option could be the use of the intergrase plasmids attempted previously without successful ligation and transformation. Transformation into X90, a more competent strain of cells according to this project, might yield better results. Additionally the use of the P_R-lacO fragments instead of the long sequence first ordered might lead to more successful ligation. This system would utilize β-galactosidase instead of GFP. This reporter gene has been thoroughly studied, and

assay of its behavior would provide an accurate picture of Cro repression. This system could also prevent the copy number issues of Cro because the whole system would integrate into the chromosome.

Another possibility could be the insertion of Cro with the control sequence into the chromosome. The P_R-lacO-GFP plasmid could then be transformed into the cell to allow for real time fluorescent assays. Any of these options could provide more reliable results than the current system.

BIBLIOGRAPHY

1. Griffiths, Anthony J.F.; Miller, Jeffrey H.; Suzuki, David T.; Lewontin, Richard C.; Gelbart, William M. An Introduction to Genetic Analysis. W.H. Freeman and Company: New York, NY. 7th ed. 2000
2. Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter. Molecular Biology of the Cell. Garland Science: New York, NY. 4th ed. 2002.
3. Albright, R.A. and B.W. Matthews. *Crystal structure of lambda-Cro bound to a consensus operator at 3.0 A resolution.* Journal of Molecular Biology: 280(1). pps 137-51.
4. Madigan, Micheal T.; Martinko, John, M.; Parker, Jack. Brock Biology of Microorganisms. Prentice Hall: Upper Saddle River, NJ. 10th ed. 2003.
5. Tsien, Roger Y. *The Green Fluorescent Protein.* Annual Reviews of Biochemistry. 67: pps 509-44. 1998.
6. Southward, Carolyn M.; Surette, Micheal G. *The dynamic microbe: green fluorescent protein brins bacteria to light.* Molecular Microbiology: 45(5). pps 1191-1196. 2002.
7. Brejc, K. and T. Sixma. *Structural basis for dual excitation and photoisomerization of the Aequorea Victoria green fluorescent protein.* PNAS: 94. p2306. 1997.
8. Kalir, S; McClure, J.; Pabbaruju, K.; Southward, C.; Ronen, M.; Leibler, S.; Surette, M.G.; Alon, U. *Ordering Genes in a Flagella Pathway by Analysis of Expression Kinetics from Living Bacteria.* Science: 292. pps. 2080-83. 15 June 2001.
9. Setty, Y.; Mayo, A.E.; Surette, M.G.; Alon, U. *Detailed map of a cis-regulatory input function.* PNAS: 100(13). pps 7702-7. 24 June 2003.
10. Ronen, Micheal; Rosenburg, R.; Shraiman, B.; Alon, U. *Assigning numbers to arrows: Parameterizing a gene regulation network by using accurate expression kinetics.* PNAS: 99(16). pps 10555-69. 6 August 2002.
11. Ptashne, Mark. A Genetic Switch: Phage lambda revisited. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. 3rd ed. 2004.
12. Ptashne, Mark. A Genetic Switch: phage λ and higher organsisms. Cell Press: Cambridge, Mass. 1992.
13. Takeda, Yoshinori; Sarai, Akinori; Rivera, Victor. *Analysis of the sequence specific interactions between Cro repressor and operator DNA by systemic base substitution experiments.* PNAS: 86. pps. 439-443. January 1989.
14. Kobiler, Oren; Rokney, A.; Friedman, N.; Court, D.; Stavans, J.; Oppenheim, A. *Quantitative kinetic analysis of the bacteriophage lambda genetic network.* PNAS: 102(12). 22 March 2005.
15. Mossing, Michael; Bowie, J.; and Sauer, R. *A Streptomycin Selection for DNA-Binding Activity.* Methods in Enzymology: 208. pps 604-19. 1991.

16. Liang, T.; Xu, C.; Dennis, P,; Bremer, H. *mRNA Composition and Control of Bacterial Gene Expression*. Journal of Bacteriology: 182(11). pps. 3037-44. June 2000.
17. Haldimann A. and Wanner B.L. *Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria.* Journal of Bacteriology: 183(21). pps. 6384-93. November 2001.
18. Ramsey, Stephen; Orrell, David; Bolouri, Hamid. *Dizzy: Stochastic Simulation of Large-Scale Genetic Regulatory Networks*. Journal of Bioinformatics and Computational Biology: 3(2). pps 415-36. 3 April 2005.