Modeling lamda Phage

Jay Raol

September 10, 2003

Abstract

We present a model of λ phage based on a synthesis of fundamental topics in biochemistry including cooperative binding, enzyme kinetics and dimerization of protiens. These ideas are finally incorporated into a set of piece wise continous differential equations. Numerical Analysis of the equations reveal that the equations do simulate the time course concentrations of the λ proteins. Hopefully the methods empolyed here will be general enough to be applied to any gene-gene interaction.

1 Introduction

λ phage is a virus which attacks Escherichia coli. In his book, "A Genetic Switch," Mark Ptashne systematically uncovers the mechanics behind the phenomenon first discovered by André Lwoff. When *E. coli* is irradiated with ultraviolet light, the bacteria stop growing and eventually lyse λ phage. These phages infect more bacteria. However, some bacteria do not lyse, but instead carry the genetic code of the phage, dormant until it itself is irradiated with UV light. Figure 2 illustrates the life cycle of the λ phage.

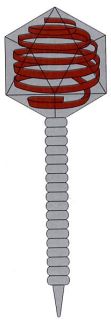

Figure 1: Structure of lambda phage containing a protein head filled with dsDNA and a protein tail used to penetrate through the bacterial cell membrane. Picture from Ptashne

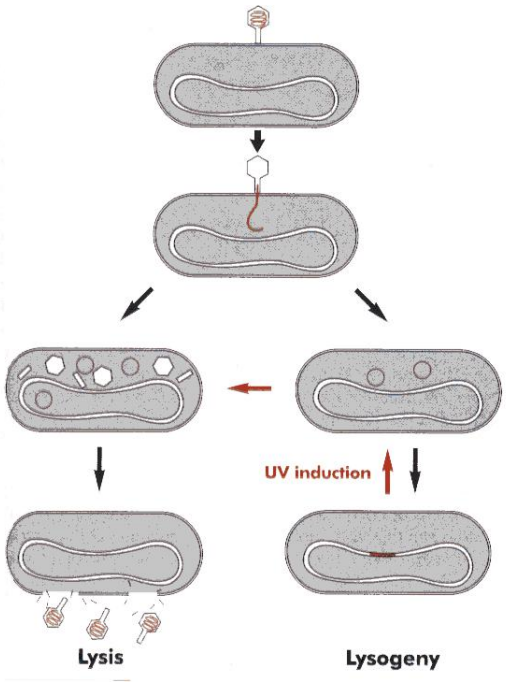

Figure 2: Illustration of the lambda phage life cycle. Picture from Ptashne

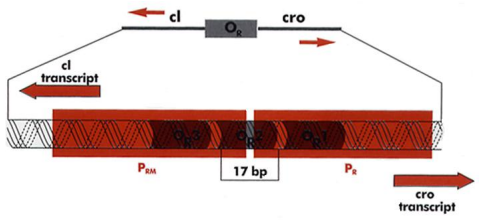

Figure 3: The cro-repressor Operon

This biological switch illustrates a fundamental property of living organisims - the ability to regulate gene expression. The decision making process of the switch occurs at the O_R operon shown in figure 3 where cI and cro both bind. Depending on the concetrations of either protein, the virus will either enter the lytic or lysogenic phase.

Upon infection, the bacteria's own RNA polymerase begins transcription of the genes which encode proteins N, cro, cIII, cII and many others. The protein N then comes back and binds the operator sites, O_L and O_R, which initiates transcription of the same genes by RNA polymerase at the promoter sites, P_L and P_R. Figure 4 shows the location of the genes of λ phage on the genome.

If the phage becomes lytic, then subsequently, the proteins Q and the dimer cro initiate transcription of all lytic genes and repress all lysogenic genes. Cro binds to its own operator site, O_R3, to ensure production of itself in a positive feedback loop and represses all other

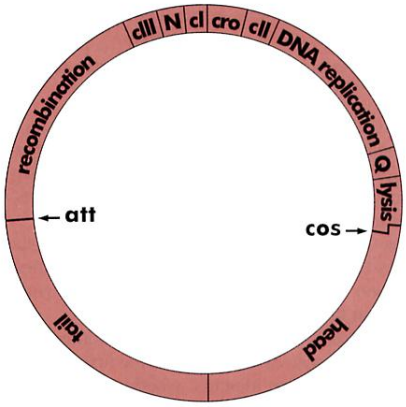

Figure 4: The lambda phage genome before insertion

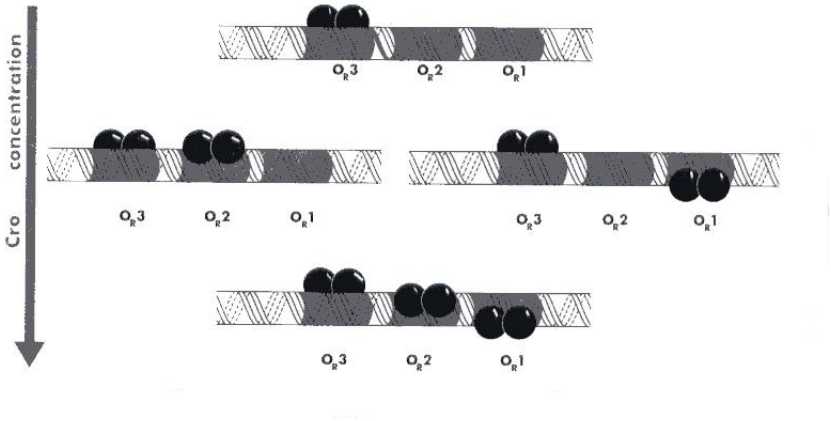

Figure 5: Cro binds to the operon site in an ordered manner due to the higher binding affinity at O_R3. As concentration increases, cro binds first to O_R3, then to either of the other two sites before finally becoming filled

genes by binding at operator site O_L. However, to ensure that its own production does not increase without bounds, λ phage has developed a special strategy. The operator site which controls transcription of cro has three binding sites. However, cro binds to each site differently. Its affinity, measured by the equilibrium disociation constant K_{cro}, is less at O_R3 than either O_R2 or O_R1, implying that $K_{cro}O_R3 \ll K_{cro}O_R2 = K_{cro}O_R1$. When cro binds to O_R3 its blocks transcription of cI. When it binds O_R2, it recruits RNA polymerase to transcribe itself. Finally, when it binds to O_R1, it blocks its own transcription. In this way, as [cro] increases, its own production rate will initially increase until enough cro causes the termination transcription. The higher affinity of cro for O_R3 ensures that the first available cro will bind to that operator site before the other two. Figure 5 shows the ordered process of cro binding to the operon site.

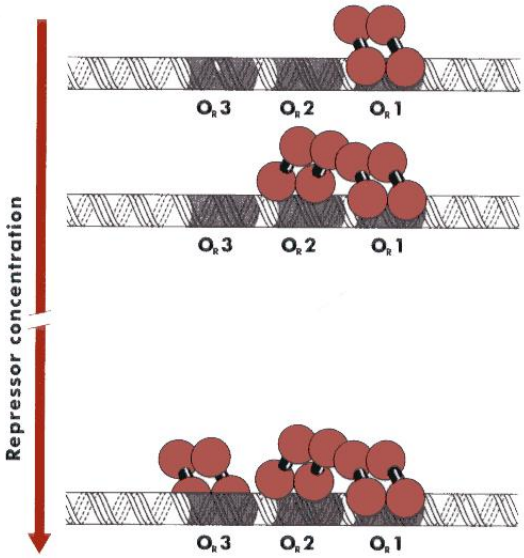

Figure 6: cI binds in a more odered fashion than cro.

Instead, if the phage becomes lysogenic, it takes another path. During the late lysogenic phase, cII protein initiates trancription of the dimer cI and Int at P_{int} and P_{RE}. Once cI is activated, it binds to O_R1 and O_L and represses all other early phase and lytic genes including cro, cII, Q, and N. It also binds O_R2 and initiates its own transcription like cro. Similarly, cI binding to O_R3 causes it to block RNA polymerase from trancribing itself. The protein cI uses cooperative binding to form this controlled positive feedback loop. Binding of either O_R1 or O_R2 induces a greater affinity at the other site. This has the affect of decreasing the overall equilibrium disociation constant. The third operator, 0_R3, does not bind cooperatively. In this way, cI will out compete cro for binding sites at lower concentrations. Therefore, we can see that at the beginning, although protein, for both the lytic and lysogenic pathways are trancribed and translated, cI will outcompete its rivals due to its greater affinity for the operator site and establish its dominance - the lysogenic pathway. Figure 6 below illustrates the order of cI binding to the O_R3 operon.

However, some phages after they have established the lysogenic pathway will lyse upon irradiation with UV light. The explanation to this lies in the concentration of cII. The protein cII by itself is not a very stable protein as it is easily susceptible to proteases in the bacteria that quickly degrade it. The protein cIII interacts with cII to protect it from degradation. Upon entering a cell, if the bacteria is healthy, it will have high levels of proteases. These proteases will affect the levels of cII and cIII. If enough is degraded, not enough cI will be produced, and the lytic pathway will take over. If the bacteria is not very healthy, there will not be enough proteases to degrade cII and the lysogenic pathway will be maintained. Another point in the pathway that is subject to outside conditions is the concentration of the cI monomer. The protein cI dimerizes with itself before it binds to DNA. When bacteria is irradiated with UV light, the UV light causes distortions of the

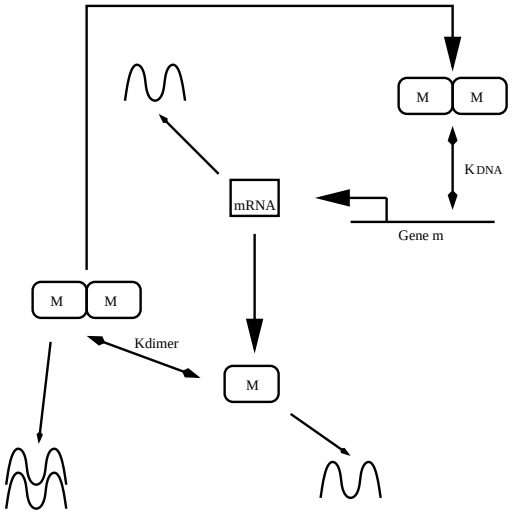

Figure 7: Half of the cro-cI λ phage switch

DNA, damaging it. In response to massive damage, the bacteria activates the SOS pathway. The SOS pathway encodes for proteins such as rec A, umu C, and uvr B, which help to repair DNA. In additon, the protein rec A cleaves the monomer cI causing the dimer cI concentrations to fall. As the cI concentration falls, it allows production of cro until the lytic pathway takes over. In this way, λ phage is able to "detect" that the cell is dying and leaves before the ship sinks.

Before we can model the following system, we must make several assumptions and adendums to the λ phage model. As proteins are produced, they associate and disossciate from DNA quickly. Our model however, only looks at concentration changes on the scale of seconds. Thefore, over a given time that is much greater than the average time of protein binding and unbinding to DNA, as protein concentrations increase, so does the average time over which the protein is bound to the DNA versus unbound to the DNA. Another assumption maintains, that they system is in equillibrium. The figure below represnts a simply positve feedback loop, where mRNA is trancribed into a protein which dimerizes with itself and binds back to its own gene. All the products are subject to degredation as shown in 1.

1.1 Dimerization

In order to explore the idea of the genetic switch, for now, we assume no degredation to explore the affects of dimerization and cooperative binding on gene expression. First, we define the relationship between the amount of monomer and the amount of dimer. Intuitively we know, that it should only depend on the equilibrium constant K_{dimer} and the total amount of monomers found in the system whether as a monomer itself or two monomers bound up into a dimer.

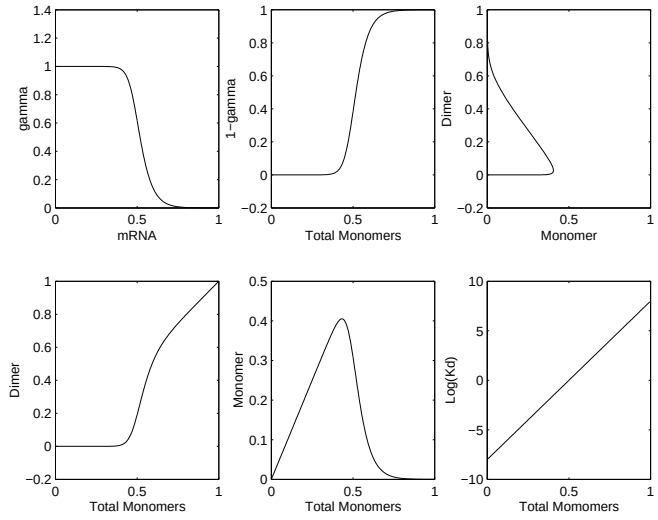

Figure 8: Plot of γ_b for varing K_{dimer} and $[M_{total}]$

$$M + M \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} M_2 \quad \text{and} \quad K_{dimer} = \frac{[M_2]}{[M]^2} \tag{1}$$

From 1, we say the total amount of monomers in the system is $[M_{total}] = [M] + 2[M_2]$ where the concentrations are given in molar. Then the ratio (γ_b) of monomers to the total amount of momomers is $\gamma_b = \frac{[M]}{[M_{total}]} = \frac{[M]}{[M]+2[M_2]}$ and the ratio of the amount of dimer to total monomers is $(1-\gamma_b) = \frac{[2[M_2]}{[M]+2[M_2]}$. Since $[M] = \gamma_b[M_{total}]$ and $[M_2] = \frac{1}{2}(1-\gamma_b)[M_{total}]$ then

$$K_{dimer} = \frac{[M_2]}{[M]^2} = \frac{\frac{1}{2}(1-\gamma_B)[M_2]}{\gamma_b^2[M]^2} \tag{2}$$

If we solve for γ_b we get a quadratic solutions, whose only valid solution is

$$\gamma_b = \frac{-1 + \sqrt{1 + 8[M_{total}]K_{dimer}}}{4[M_{total}]K_{dimer}} \tag{3}$$

The graph of this γ_b as the total amount of monomers increases follows a hyperbolic trajectory. K_{dimer} controls the amount of dimer and monomer produced. A strong dimerization affinity, $K_{dimer} = 10^{-8}$ greatly favors dimer production. 1.1 is a graph of γ_b at varying levels of affinity.

1.2 Enzyme Kinetics and Saturation Curves

Now, that we can describe the ratio of monomers and dimers to the total amount of momomer subunits, lets us proceed to describe the cooperative binding of a ligand to DNA. Although DNA can not be considered a concentration in the normal sense, we can use the term "[DNA]" to help describe the binding of ligand D to DNA. Then we can use saturation curves with no dependance on any concentration of DNA to interpret the binding of the ligand to DNA. If Y is the ratio of ligand bound DNA to total DNA, then Y equals

$$Y = \frac{[DNA - D]}{[DNA - D] + [DNA]} \quad \text{and} \quad K_{DNA} = \frac{[D]^n \ [DNA]^n}{[DNA - D]^n} \tag{4}$$

Substituting the definiton of the disassociation constant K_{DNA} into the expression for Y gives

$$Y = \frac{[D]^n}{[D]^n + K_{DNA}} \tag{5}$$

In order to describe more complex regulation at the recruitment sites for RNA polymerase, we turn to enzyme kinematics, specifically how the binding of an inhibitor effects the velocity and half-saturation.

One complication to this methos is cooperative binding. Cooperative binding describes the effect where the binding of one ligand can enhance the binding of subsequent ligands. Usually, after the first ligand binds to its complex, it causes some conformational change wich allows the next ligand to bind more or less easily. The study of hemoglobin offers a great example of this phenomenon and a solution on how best to model it. Hemoglobin binds four oxygens in the lungs and carrys it to the cells. Classicaly, we describe this equation as $H + 4O_2 \rightleftharpoons HO_2^4$ and $K_d = \frac{[H][O_2]^4}{[HO_2^4]}$. However, experiments show that oxygen binding to hemoglobin does not have a fourth power dependance as the equilibrium constant for this reaction would predict. Instead, n varies depending on how many oxygen are already bound to hemoglobin. One way to get empirically solve this system, is to rewrite the equation to isolate n. Using saturation, Y, we can derive the Hill equation

$$\log\left(\frac{Y}{Y-1}\right) = n \log O_2 + \log K_d \tag{6}$$

However, instead of this approach, there is another approach to describing cooperative binding. Looking at the figure below, we can see that although the same ligand binds twice to DNA, it has to different binding constants. Therefore, we can treat the system as two different species binding with two different equilibrium constants.

We can combine the methods of enzyme kinetics and cooperative binding to fully describe the cro-cI system in λ phage. Wwe assume that the time course over which we observe Y is much greater than the time it takes for the association and disassociation of the RNA

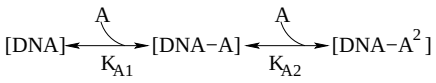

Figure 9: The the series of ordered reactions that describe cooperative binding

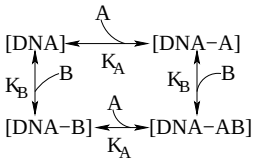

Figure 10: A system where ligands B and A bind to DNA in a non-ordered fashion

transcriptional machinary, then we can say, Y simply gives the fraction when D was bound to DNA at a given time t. In additon, we can also say that there is some validity to calling "[DNA-D]" instead of a single molecule. The strength of Y is in its ability to weight by the binding affinity of the protein in determining which proteins will determine wiether or not RNA polymerase will be recruited. The following set of chemical equations describe the binding of proteins A and proteins B. Protein A recruits RNA polymerase and protein B represses it whether A is bound or not. Finally, the saturation (Y) is simply the ratio of all posible promoter to all possible species.

$$\begin{aligned} [DNA] + [A] &\overset{K_A}{\rightleftharpoons} [DNA - A] \\ [DNA] + [B] &\overset{K_B}{\rightleftharpoons} [DNA - B] \\ [DNA - A] + [B] &\overset{K_B}{\rightleftharpoons} [DNA - A - B] \\ [DNA - B] + [A] &\overset{K_A}{\rightleftharpoons} [DNA - B - A] \end{aligned} \tag{7}$$

$$Y = \frac{[DNA - A]}{[DNA - A] + [DNA - B] + [DNA - A - B] + [DNA]} = \frac{\frac{[A]}{K_A}}{\frac{[A]}{K_A} + \frac{[B]}{K_B} + \frac{[A][B]}{K_A K_B} + 1} \tag{8}$$

$$Y = \frac{\frac{[A]}{K_A} + \frac{[B]}{K_B} + \frac{[A][B]}{K_A K_B}}{\frac{[A]}{K_A} + \frac{[B]}{K_B} + \frac{[A][B]}{K_A K_B} + 1} \tag{9}$$

Figure 11 shows the satuation for various K_A and K_B values at two different concentrations of A and B. The top two graphs are plots of 8 and the bottom two of 9. The most noticeable aspect of these graphs is the sensitivity of the saturation to variations in concentrations and affinities. The saturation is able to stay near 1% of total saturation for a wide range before the saturation drops steeply off to zero.

From the general principles outlines thus far we can abstract a more general way of qualifying Y. We have thus far defined a system where two ligands, A and B, have bound to DNA in different logical ways. Both A and B are necessary for transcription or either A or B is enough for transcription. We can define the logic operator not as well. For a system where the DNA alone will cause trancription and the binding of protein A will inhibit it, we can describe the saturation as the following

$$Y = \frac{1}{\frac{[A]}{K_A} + 1} \tag{10}$$

Equation 10 shows that the relationship between A as a promotor and A as a repressor is simply substractive. We can similarily describe the logical "and" and "or". If A and B represent Michaelis-Menton like terms defined by $A = \frac{[a]}{[a]+K_a}$ where K_a is the equillibrium disassociation constant. then

$$\begin{aligned} \text{not} \quad & A = 1 - A \\ A \quad \text{and} \quad & B = A * B \\ A \quad \text{or} \quad & B = A + B(1 - A) \end{aligned} \tag{11}$$

In this way, the saturation term (Y) encodes the logic at the regulation sites in a conitous manner. However, this allows the recruitment of RNA polymerase to take on not only "on" or "off" values but also intermediate values. The binding affinities of the individual proteins determine to what degree the gene is "on" or "off". If the affinity is very high, it will take only a small amount to saturate the system, and we consider that it interacts with the system in a boolean fashion. However, if the affinity is low, then the saturation of the system is dependant on the contious concentration of the protein.

In to quantify the relationship between the production of mRNA and gene operon saturation, we assume that if a promoter is covelantly linked to DNA, then RNA is trancribed at a terminal rate of α. We can then say that the rate of transcription of mRNA is αY.

1.3 Piece-wise ODE's

Using 3 and 5, we can now fully describe the system in Figure 1. In addition, we will now consider a system where the monomer, dimer, and mRNA all get degraded. For now we will only model the constitutive degredation pathway. Therefore, all molecules in the cell will degrade at a rate proportional to their concentration. However, as seen in many experimental results, we assume mRNA degrades very rapidly after a give time, A. Our system of equations then becomes

$$\frac{d[mRNA]}{dt} = \alpha Y - \varepsilon[mRNA] \tag{12}$$

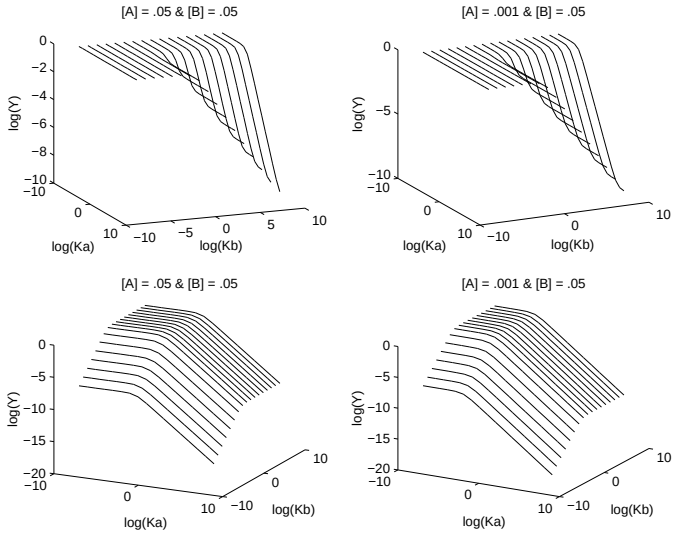

Figure 11: Saturation Curves for varying affinities, concentrations, and logic

$$\begin{aligned} Y &= \frac{[D]^n}{[D]^n + K_{DNA}} \\ \frac{d[M]}{dt} &= \beta[mRNA] + k_{-1}[D] - (k_1 + \varepsilon)[M] \\ \frac{d[D]}{dt} &= k_1[M] - (k_{-1} + \varepsilon)[D] \end{aligned}$$

Another factor to consider are the degredation and production terms α, β and ε. These are infact maximum possible values. We use them here as convienent approximations. The production and degredation rates all follow a Michaelis-Menton type interation since we are discussing enzyme reactions. Therefore, we can write explicity value of these terms.

$$velocity = \frac{\text{Maximal rate}}{[substrate] + K_{affinity}} \tag{13}$$

If since we only incoporated the constitutive degredation pathway into our model, we can say that each species begin degraded have equal affinity for the proteosomes. Rewriting 12, our new set of equations becomes

$$\begin{aligned} \frac{d[mRNA]}{dt} &= \alpha Y - \frac{\varepsilon}{[mRNA] + K_{deg}} \\ Y &= \frac{[D]^n}{[D]^n + K_{DNA}} \\ \frac{d[M]}{dt} &= \frac{\beta}{[mRNA] + K_{prod}} + k_{-1}[D] - k_1[M] + \frac{\varepsilon}{[M] + K_{deg}} \end{aligned} \tag{14}$$

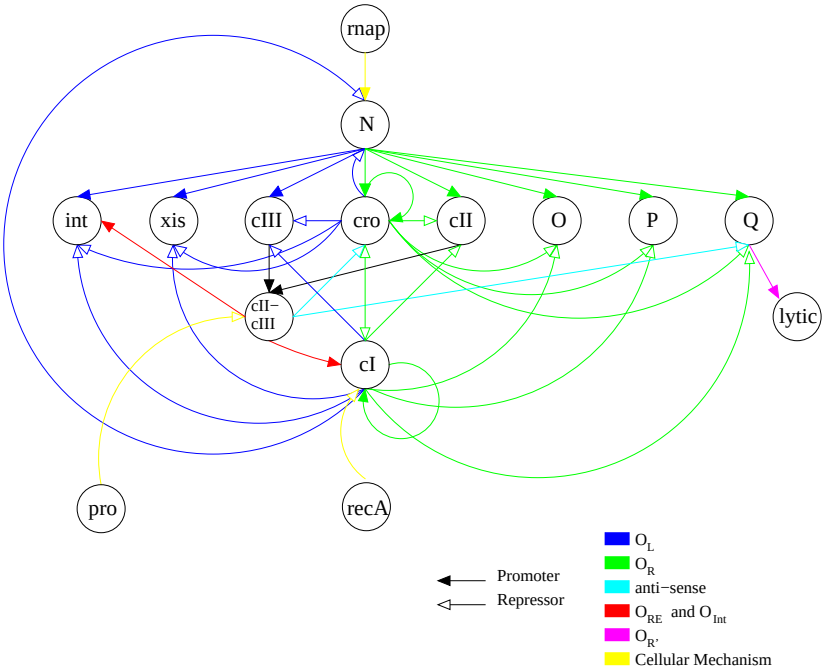

Figure 12: Connectivity graph for the phage system. The sepeate colors represent the operon upon which the ligand will act. In this way, we can seperate the mode of activation/repression of different pathways.

$$\frac{d[D]}{dt} = k_1[M] - k_{-1}[D] + \frac{\varepsilon}{[D] + K_{deg}}$$

The following to figure are the plots of the solutions of 12 and 14 where several different starting conditions are graphed for each equation set.

1.4 The Full system

In order to graph the full λ phage system, we must couple to genes that inhibit each other and damped positively excite themselves. The cro gene will bind to DNA non-cooperatively, while cI will bind in such a cooperative manner. The other genes in the system bind non-cooperatively in a more strait forward regulatory way with very few species invloved.

We now have the ability to quantify the following connectivity graph with the following set of ODE's. Most genes will be set up exactly as equation 12. The major difference will be in the saturation terms. The following equations are the saturation terms which describe the connectivity graph in 12.

$$
\begin{aligned}
Y_N &= \frac{\frac{[N]}{K_N}+1}{\frac{[N]}{K_N}+\frac{[cI]}{K_{cI}}+\frac{[cro]}{K_{cro_N}}+1} \qquad (15)\\
Y_{cro} &= \frac{\frac{[cro]^2}{K_{cro1}K_{cro2}}+\frac{[N]}{K_N}}{\frac{[cro]^2}{K_{cro1}K_{cro2}}+\frac{[N]}{K_N}+\frac{[cro]^2}{K_{cro1}K_{cro2}}+\frac{[cro]^3}{K_{cro1}K_{cro2}^2}+\frac{[cI]^2}{K_{cI1}}+\frac{[cro]^2[cI]}{K_{cro1}K_{cro2}K_{cI2}}+\frac{[cI]^3}{K_{cI1}K_{cI2}}+1}\\
Y_1 &= \frac{\frac{[N]}{K_N}}{\frac{[N]}{K_N}+\frac{[cI]}{K_{cI}}+\frac{[cro]}{K_{cro_N}}+1}\\
Y_{lytic} &= \frac{[Q]}{[Q]+K_Q}\\
Y_{cI} &= \frac{\frac{[cI]^2}{K_{cro1}}+\frac{[N]}{K_N}}{\frac{[cro]^2}{K_{cro1}K_{cro2}}+\frac{[N]}{K_N}+\frac{[cro]^2}{K_{cro1}K_{cro2}}+\frac{[cro]^3}{K_{cro1}K_{cro2}^2}+\frac{[cI]^2}{K_{cI1}}+\frac{[cro]^2[cI]}{K_{cro1}K_{cro2}K_{cI2}}+\frac{[cI]^3}{K_{cI1}K_{cI2}}+1}\\
Y_{cII_cIII} &= \frac{[cII]}{cII+K_{cII}}\\
Y_{cII} &= = \frac{\frac{[cII]}{K_{cII}}}{\frac{[cII]}{K_{cII}}+\frac{[cro]}{K_{cro}}+\frac{[cI]}{K_{cI}}+1}
\end{aligned}
$$

Notice in the expression for Y_N that addition of 1 in the numerator signifies that the protein N can be transcribed by RNAP without any promoters at all. In addtion, Y_1 represents all the generic lytic and lysogenic protein including xis, O, P, Q, and cII who share the same regulation. The network in figure 12 can be described by the following set of ODE's

$$
\begin{aligned}
\frac{d[mRNA]_N}{dt} &= \alpha Y_N - \varepsilon[mRNA]_N \qquad (16)\\
\frac{d[N]}{dt} &= \beta[mRNA]_N - \varepsilon[N]\\
\frac{d[mRNA]_{xis,O,P}}{dt} &= \alpha Y_1 - varepsilon[mRNA]_{xis,O,P}\\
\frac{d[xis,O,P]}{dt} &= \beta[mRNA]_{xis,O,P} - \varepsilon[xis,O,P]\\
\frac{d[mRNA]_{Int}}{dt} &= \alpha(Y_{Int}+Y_{cII_cIII})+\varepsilon[mRNA]_{Int}\\
\frac{d[Int]}{dt} &= \beta[mRNA]_{Int} - \varepsilon[Int]\\
\frac{d[mRNA]_Q}{dt} &= \alpha Y_1 - (\varepsilon + K_{antiQ}[ssimRNA]_Q)[mRNA]_Q\\
\frac{d[Q]}{dt} &= \beta[mRNA]_Q - \varepsilon[Q]
\end{aligned}
$$