Catching pUC19 Red-Handed

Creating Reporter Constructs to Analyze Gene Expression

Ross Lancaster

Biology 4270

Dr. Rutledge

11/1/20

Introduction:

In the world of biotechnology, it is very common to want to analyze gene expression within a cell. This allows one to study such topics as cellular differentiation, protein protein interactions, and metabolism. One easy way to accomplish this is to use ***reporter genes.*** Reporter genes have easily identifiable products, such as those that are fluorescent. Two common reporter genes are Green Fluorescent Protein (GFP) and Red Fluorescent Protein (RFP). These genes add ***protein tags*** as they code for proteins that glow bright green and red, respectively. If the reporter gene is placed directly in front of the gene of interest, GFP and RFP add a glowing tag to the protein being studied. This allows one to see when and where the gene of interest is being expressed. The other common type of protein tag is an affinity tag. This is where the protein tag is recognizable by another molecule, often an antibody, which can bind to the tag and detect the protein (Dale 2012).The reporter gene can then be attached to the promoter of the gene of interest to create a reporter construct. Thus, once the promoter is activated in the cell, both the reporter gene and the gene of interest are expressed. This allows one to study the activity of the promoter through the product of the reporter gene. (Dale, 2012).

To fully understand reporter genes, it is important to have a strong grasp on bacterial gene expression. The central dogma states that DNA is transcribed into RNA which is translated into protein. Figure 1 depicts the central dogma and indicates the promoter region with a black arrow and RNA polymerase binding to the promoter with a red arrow. The transcription of many bacterial genes is controlled by ***operators.*** Operators are regions of DNA that proteins can bind to and affect gene expression. (Dale 2012). Proteins bind to DNA

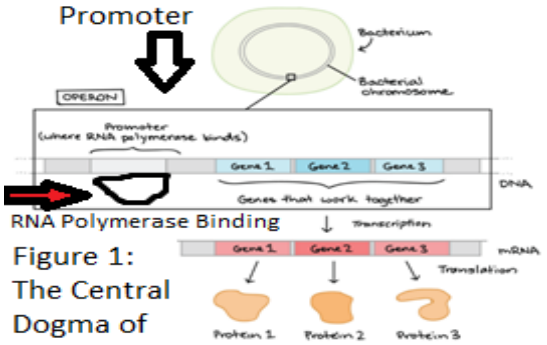

Figure 1: The Central Dogma of Biology States that DNA is Transcribed into RNA and Translated into Protein[1]

through multiple ***motifs***, or substructures of a protein whose specific shape make binding to the DNA favorable. These substructures interact with the DNA through hydrogen bonds and van der Waals interactions. They often interact with the major groove. One of the most common motifs is the helix-turn helix motif (Harrison, 1991). The black arrow in Figure 2 exhibits this motif interacting with the major groove of a DNA strand. Once bound, proteins can make structural changes to the DNA to affect gene expression. For example, araC, an important regulatory protein, affects gene expression with helix-turn-helix motifs on its dimers. These bind to two separate spots on the DNA strand. The dimers then bind together to cause DNA looping that physically prevents RNA polymerase from binding (Schleif, 2003).

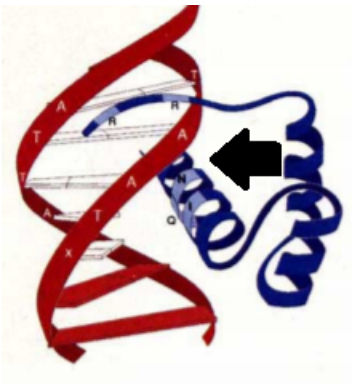

Figure 2: Helix-Turn-Helix Motif Interacting with the Major Groove of DNA[2]

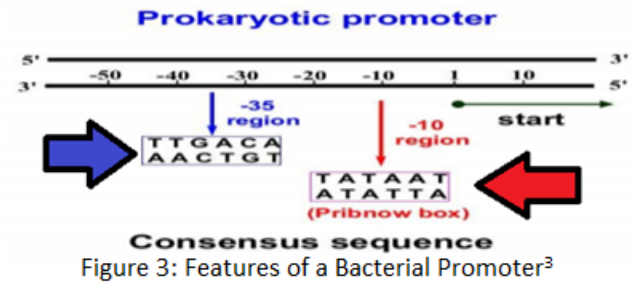

Figure 3: Features of a Bacterial Promoter[3]

Another common spot where proteins bind to DNA strands is the ***promoter.*** The promoter is the region of DNA upstream of the start of transcription where RNA polymerase binds. Sigma factors can act as accessory proteins to RNA polymerase to add even more specificity to promoter binding (Dale, 2012). After binding to the promoter RNA polymerase is able to create an mRNA transcript that will be translated into a protein. Figure 3 depicts the common features of the bacterial promoter; the blue arrows points to the ***-35 box***, which is 35 base pairs upstream of the start of transcription, while the red arrow points to the ***Pribnow box***, which is 10 base pairs upstream of the start of transcription. Both the -35 box and the Pribnow box serve as ***consensus sequences*** that subunits of RNA polymerase bind to. The -35 box has a consensus sequence of TTGACA while the Pribnow box has the

consensus sequence TATAAT. Not all bacterial promoters utilize these exact sequences. However, in any bacterial promoter every base in the -35 and Pribnow boxes is likely going to be same base as the one found in the consensus sequence. (Dale, 2012).

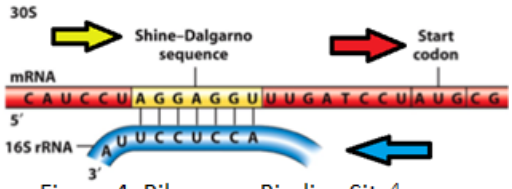

Figure 4: Ribosome Binding Site[4]

Once RNA polymerase has created an mRNA transcript, this transcript needs to be translated into a functional protein. mRNA transcripts are translated by ribosomes, which are able to bind to the mRNA transcript through the ***ribosome binding site***. In bacteria this is known as the Shine-Dalgarno sequence (Dale, 2012). In Figure 4, the yellow arrow points to the Shine-Dalgarno sequence while the blue arrow shows the ribosome binding to the DNA strand upstream of the start codon, shown by the red arrow. The ribosome binds to this purine rich region and moves downstream until it hits a start codon. Another important aspect of translation is the ***reading frame.*** The reading frame starts with the start codon and ends with the stop codon. However, because codons are made of three nucleotides it is important that the ribosome starts translating at the correct nucleotide. Each double stranded DNA molecule has six reading frames. The Shine-Dalgarno sequence helps the ribosome find the correct reading frame as it guides the ribosome to bind just upstream of the correct start codon. Using the wrong reading frame would create a completely different or nonfunctional protein.

The complete central dogma can be seen through the creation of the β-lactamase protein (Figure 5). β-Lactamase is important because many reporter constructs will contain the ampicillin resistance gene. Thus, when plated on a medium with ampicillin, any bacteria without the reporter construct will be killed by ampicillin. However, those with the reporter construct will express the ampicillin resistance gene and express β-Lactamase. β-lactamase then

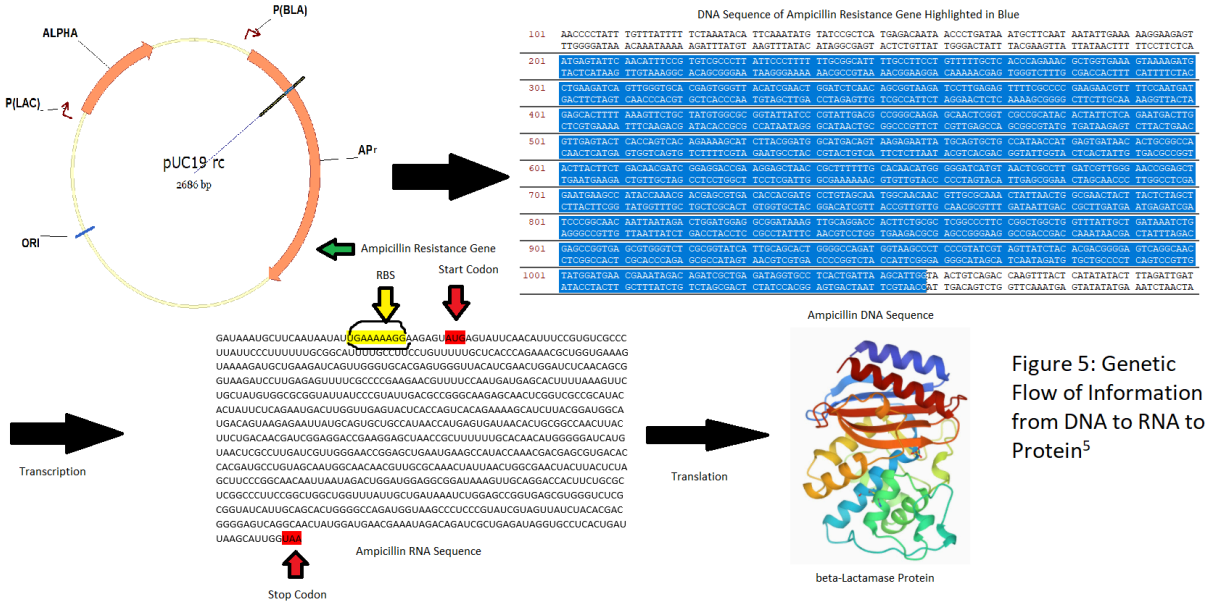

Figure 5: Genetic Flow of Information from DNA to RNA to Protein[5]

hydrolyzes ampicillin to render it inactive. Figure 5 starts with the pUC19 plasmid and then shows the DNA sequence of the ampicillin resistance gene. This sequence is then transcribed into the mRNA transcript which is then translated by the black ribosome on the RBS. Finally, one sees how the final β-Lactamase protein folds into its specific structure.

Results:

Two main reporter constructs will be reviewed, RFP in the pUC19 plasmid and the pGLO plasmid. Both of these plasmids have genes that code for fluorescent proteins attached to a promoter. Bacteria are then induced to take up these plasmids. Bacterial cells that activate the promoter are then visually distinguishable by the fluorescent protein (Dale, 2012).

RFP in pUC19 is used to study the lac operon promoter. The major components of pUC19 are the lac operon promoter, the alpha subunit, and the ampicillin resistance gene (Figure 6). In order to study the lac operon

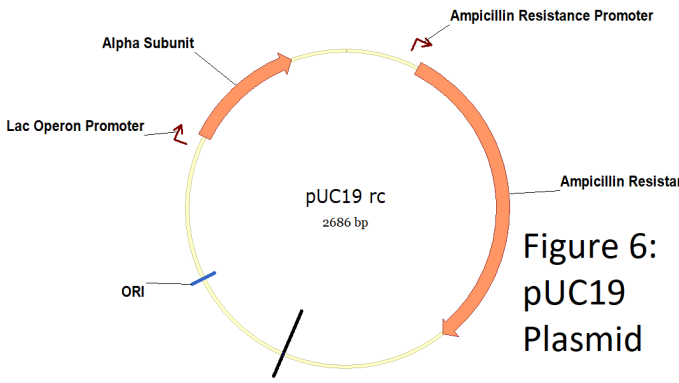

Figure 6: pUC19 Plasmid

promoter the RFP gene will need to be added to the plasmid. The first step in the process is to create a copy of the RFP gene on another plasmid through PCR. To create a copy of a gene DNA polymerase needs an existing strand of nucleotides to extend. Primers are specifically designed DNA sequences that provide these nucleotides by annealing to the plasmid with RFP. The primers are then extended by DNA Polymerase to copy RFP. Figure 7 shows the primer sequences that were used to make a copy of the RFP gene on the coel:RFP plasmid. The yellow

Primer Sequences:

Forward Primer: 5′-GTCGGATCCCATGGTGCGCTCCTCCAAGAA-3′ Yellow=Random Nucleotides

Reverse Primer: 5′-GCATGGTACCCTACAGGAACAGGTGGTGGC-3′ Green=Cut Site Red= Reading Frame

Figure 7: Primer Sequences

bases are random and are added because restriction enzymes need bases flanking the cut sites to bind and cut accurately. The green bases are the cut sites for the restriction enzymes. The red base was a random base added to maintain the reading frame of pUC19. Figure 8 shows how the reading frame of pUC19 was maintained. RFP begins with the amino acid sequence MVRSSK. If the nucleotide sequence for RFP is added directly after the BamHI cut site, then RFP is out of frame as the sequence VCAPPR is coded for. However, adding the random nucleotide to the primer maintains the reading frame as the sequence MVRSSK is still coded for.

After designing the primers, pUC19 and the copies of RFP made through PCR are digested with the same restriction enzymes. The digested RFP and pUC19 are then ligated together. Bacteria can then be transformed with pUC19 by placing them on ice and then heat

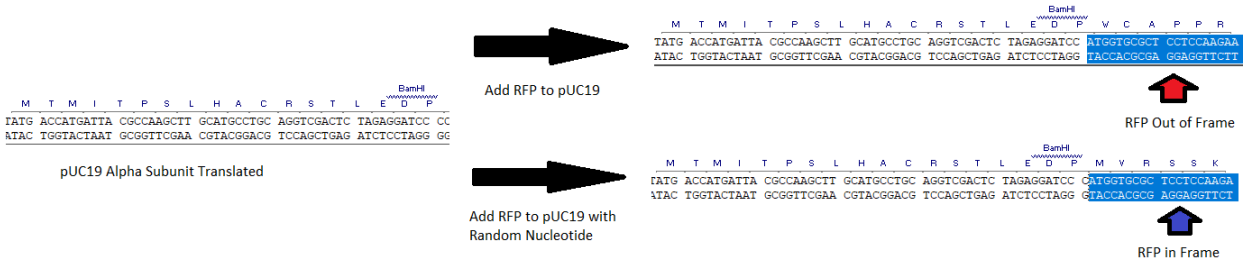

Figure 8: Maintaining the Reading Frame of pUC19

shocking them. This creates pores in the cell wall that allow the plasmid to enter the bacteria. The bacteria are then plated with ampicillin. As pUC19 contains the ampicillin resistance gene, any bacteria that do not take up a plasmid will be killed by ampicillin. Bacteria transformed with a plasmid that was successfully ligated with RFP will glow red while bacteria that cannot create functional RFP glow white. Figure 9 shows the pelleted white and red colonies under UV light. The red glow is a sign that the lac operon is being expressed. Thus, the lac operon promoter is active and able to bind RNA polymerase and transcribe RFP. Thus, RFP allowed lac operon promoter activity to be analyzed.

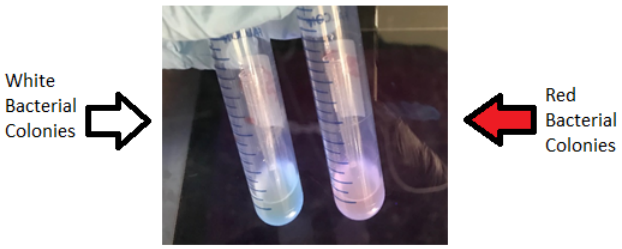

Figure 9: White and Red Bacterial Colonies Under UV Light

Another important reporter construct is pGLO. pGLO is used to analyze the pBAD promoter. The pBAD promoter regulates genes that are involved in the breakdown of arabinose. Like pUC19, pGLO has the gene for ampicillin resistance as well as a reporter gene, in this case GFP (Figure 10). GFP codes for a protein that glows green when expressed. Using GFP as a reporter gene allows one to study both where and when a promoter is active as well as the movement of signaling proteins (Tavaré, 2001). pGLO also has the AraC gene. AraC codes for a protein that represses the pBAD operon in the absence of arabinose by binding to the operators O_1 and O_2. In the presence of arabinose the AraC protein binds to the I_1 and I_2 operator regions and activates transcription of the pBAD operon. (Invitrogen, 2010). GFP has been placed next to the pBAD promoter just like RFP was placed next to the

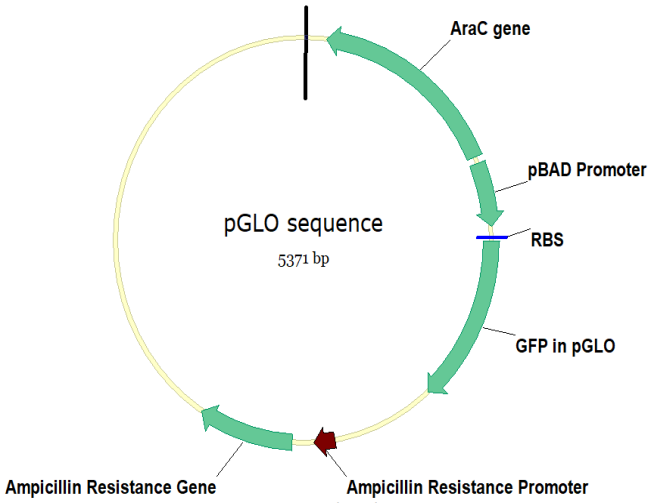

Figure 10: pGLO Plasmid

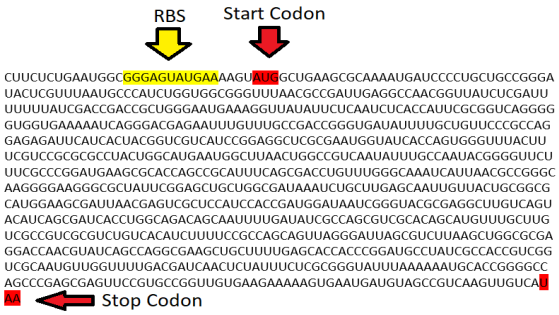

Figure 12: AraC mRNA Sequence

lac operon promoter. Thus, in the presence of arabinose the AraC gene will stimulate RNA polymerase binding to the promoter and transcription of GFP. The green glowing protein will then indicate pBAD promoter activity.

To analyze the pBAD promoter, bacteria must be transformed with pGLO and grown on a plate with arabinose. As arabinose is present, the AraC protein will be transcribed and stimulate RNA polymerase to bind to the pBAD promoter. Figure 12 shows the mRNA transcript for the AraC gene with its RBS while Figure 11 shows both the AraC and GFP genes, their translation, and their proximity to the pBAD promoter. RNA polymerase will then transcribe GFP and a ribosome will bind to the Shine-Dalgarno sequence (highlighted in blue in Figure 11). The ribosome will then translate the GFP mRNA into a glowing protein. The presence of the green protein is a visual indicator that the pBAD promoter is active. Thus, GFP allowed

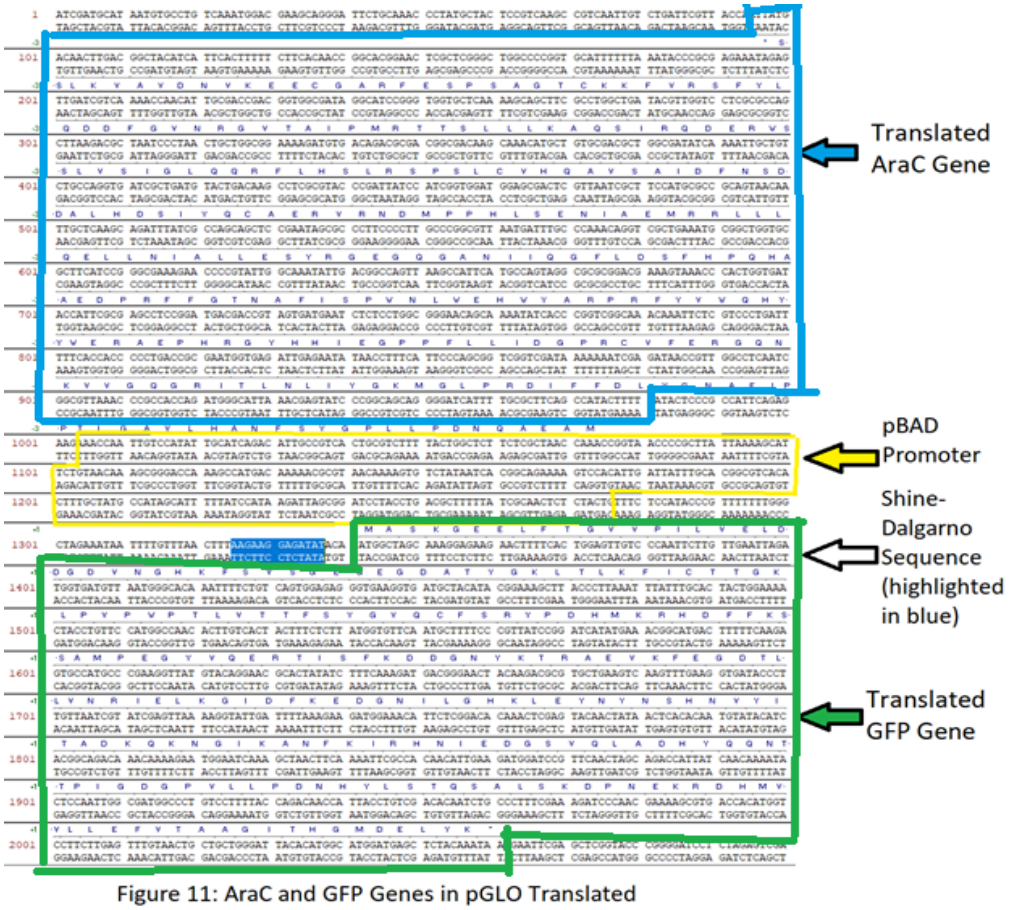

Figure 11: AraC and GFP Genes in pGLO Translated

for pBAD promoter activity to be analyzed.

Summary and Discussion:

There are many processes happening at the molecular level for transformed bacteria to glow red or green. Plasmids are transcribed by the machinery inherent to the bacteria and do not require foreign proteins. For RFP and pUC19, the lac operon must be active for the bacteria to glow red. The lac operon is a set of genes, lacZ, Y, and A, that code for proteins that break down lactose. They are all controlled by one promoter and one operator. One of the products of lactose breakdown is allolactose. Allolactose can then bind to the repressor protein, lacI. This prevents the repressor from binding to the operator sequence and blocking RNA polymerase from transcribing the Lac genes (Ralston, 2008). Thus, allosteric regulation is involved with the lac operon as allolactose binds to lacI at a site other than its DNA binding site. This causes a conformation change in lacI that prevents it from binding to the operator sequence. Thus, if the RFP gene is attached to the lac operon one can make bacteria glow red by preventing lacI from binding to the operator. Allolactose binds to lacI, allowing RNA polymerase to bind to the lac promoter and travel down the DNA to transcribe RFP. However, even in the absence of lactose RFP can be transcribed. LacI only binds to the promoter transiently. This allows for the promoter to not always be blocked and for RNA polymerase to bind and express RFP at lower levels even in the absence of lactose. A ribosome then binds to the RFP mRNA transcript and translates it into the RFP protein. Finally, the RFP protein glows red, showing that the lac promoter is active.

A similar process is happening on the molecular level when bacteria with pGLO are glowing green. The pBAD operon is regulated by the AraC protein and must be active for the bacteria to glow. AraC is a protein dimer and its N-terminal arms determine where it binds on the DNA sequence. In the absence of arabinose, the N-terminal arms bind to the DNA binding

domains of the AraC protein. This makes it more favorable for the DNA binding domains of the dimers to bind to the I_1 and O_2 regions of the pBAD operon. These regions are distant and AraC binding to them creates a loop in the DNA that physically blocks RNA polymerase from binding to the promoter. However, when arabinose is present the N-terminal arms prefer to bind to arabinose. This allows the DNA binding domains to preferentially bind to the I_1 and I_2 sites of the promoter. This position promotes RNA polymerase binding to the pBAD promoter and activates transcription. This is known as the light switch mechanism as the N-terminal arms being in either the on or off position regulates the pBAD promoter (Schleif, 2003). This is another example of allosteric regulation as arabinose acts as an allosteric effector. Arabinose binding outside of the DNA binding domain causes a conformation change in the AraC protein. This change in conformation then affects where AraC is more likely to bind. Thus, if the GFP gene is attached to the pBAD operon one can make bacteria glow green by having arabinose bind to AraC. Arabinose binding to AraC allows AraC to bind to the DNA in a way that is favorable for RNA polymerase to bind. This allows RNA polymerase to transcribe GFP. However, if a different such sugar like glucose was present, GFP would not be able to be transcribed as AraC would not undergo the conformational change necessary to activate the promoter. A ribosome then translates the mRNA transcript and assembles the GFP protein. A fluorophore within the hydrophobic core of GFP then allows the protein to glow green. RFP is capable of glowing in a similar manner as it is a mutant form of GFP (Tavaré, 2001).

It is interesting to note that repression of the lac operon and pBAD promoter is very similar. Both are repressed by the presence of a repressor protein binding to the operator sequence of the operon. Repression is also allosterically regulated for both. Allolactose binding to lacI allosterically inhibits the lac operon while arabinose binding to AraC allosterically

inhibits the pBAD promoter. Also, as both operons code for genes that break down sugars, they are also both repressed in the absence of said sugar. The lac operon codes for genes that breakdown lactose and when no lactose is present allolactose will not bind to lacI and the lac operon will be repressed. The pBAD operon codes for genes that breakdown arabinose and in the absence of arabinose the AraC protein will bind in a manner that physically prevents the operon from being transcribed.

pUC19 can also be analyzed to see if it has any of the operator DNA from the lac operon. The lac operon has three operator sequences, O_1, O_2, and O_3. O_1 is the main operator and has high affinity for lacI. O_2 and O_3 have lower affinity for lacI and are more distant from the promoter. LacI is able to bind to two of the three domains to form a loop in the DNA similar to that caused by the AraC protein to physically block RNA polymerase from binding (Oehler, 1990). Figure 13 shows the comparison of DNA sequences of the lac operon and pUC19. It shows that pUC19 contains the exact sequences for the O_1 and O_3 operators but does not contain the sequence for the O_2 operator. As lacI only needs to bind to two of the operators to repress the lac operon, Figure 3 exhibits how pUC19 has sufficient DNA sequences for the lac operator to work.

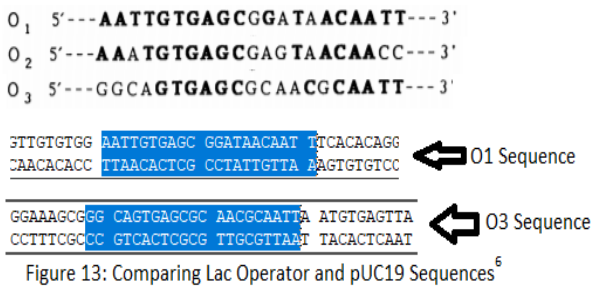

Figure 13: Comparing Lac Operator and pUC19 Sequences[6]

Finally, pGLO and RFP in pUC19 are both reporter constructs but pGLO is a transcriptional fusion vector while RFP in pUC19 is a translational fusion vector. Transcriptional fusion vectors are only provided a promoter by the host vector and provide their own translation signals (RBS and start codon) whilst translational fusion vectors use the host's translation signals (Dale, 2012). When GFP is added to pGLO it is added with its own start codon and RBS and only needs the vector's promoter region to be transcribed and translated. On the other hand,

when RFP is inserted into pUC19 the start codon and RBS for the lac genes of the lac operon are needed to transcribe and translate RFP. However, even if the translation signals bacteria need to express RFP in pUC19 and GFP in pGLO differ, they are both very useful reporter constructs for analyzing gene expression and promoter activity.

Works Cited

"From Genes to Genomes, Analysis of Gene Expression, and Products from Native and Manipulated Cloned Genes." *From Genes to Genomes: Concepts and Applications of DNA Technology*, by Jeremy W. Dale et al., John Wiley & Sons, 2012.

Harrison, Stephen C. "A Structural Taxonomy of DNA-Binding Domains." *Nature*, vol. 353, no. 6346, 1991, pp. 715–719., doi:10.1038/353715a0.

Invitrogen. "PBAD/His A, B, and C." 29 Dec. 2010. Version J

Oehler, S., et al. "The Three Operators of the Lac Operon Cooperate in Repression." *The EMBO Journal*, vol. 9, no. 4, 1990, pp. 973–979., doi:10.1002/j.1460-2075.1990.tb08199.x.

Ralston, Amy. "Operons and Prokaryotic Gene Regulation." *Nature News*, Nature Publishing Group, 2008, www.nature.com/scitable/topicpage/operons-and-prokaryotic-gene-regulation-992/.

Schleif, Robert. "AraC Protein: A Love-Hate Relationship." *BioEssays*, vol. 25, no. 3, 2003, pp. 274–282., doi:10.1002/bies.10237.

Tavare, J M, et al. "Using Green Fluorescent Protein to Study Intracellular Signalling." *Journal of Endocrinology*, vol. 170, no. 2, 2001, pp. 297–306., doi:10.1677/joe.0.1700297.

Image Citations

[1]"Overview: Gene Regulation in Bacteria (Article)." *Khan Academy*, Khan Academy, www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/regulation-of-gene-expression-and-cell-specialization/a/overview-gene-regulation-in-bacteria.

[2] Harrison, Stephen C. "A Structural Taxonomy of DNA-Binding Domains." *Nature*, vol. 353, no. 6346, 1991, pp. 715–719., doi:10.1038/353715a0.

[3] Pathshala. "Molecular Biology, Genetic Engineering, and Biotechnology: Bacterial Transformation." *Pathshala*, epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000002BI/P001357/M021478/ET/1501754242E-TextModule7Bacterialtranscription.pdf.

[4] Dnaofbioscience. "What Is Shine Dalgarno Sequence?" *Bits and Bytes of Biology*, 1 Jan. 1970, dnaofbioscience.blogspot.com/2016/05/what-is-shine-dalgarno-sequence.html.

[5]Bank, RCSB Protein Data. “1XPB: STRUCTURE OF BETA-LACTAMASE TEM1.” *RCSB PDB*, National Science Foundation, www.rcsb.org/structure/1XPB.

[6]Oehler, S., et al. “The Three Operators of the Lac Operon Cooperate in Repression.” *The EMBO Journal*, vol. 9, no. 4, 1990, pp. 973–979., doi:10.1002/j.1460-2075.1990.tb08199.x.