

# Exercise Eight

## Recombinant DNA Techniques and Their Applications: Genetic Engineering for Biofuel Production<sup>1</sup>

### In this lab you will:

- Use selection techniques (colony color, antibiotic resistance, and reporter gene expression) to determine the presence and identity of cloned genes in *E. coli* cells.
- Isolate pLOI297, a recombinant pUC18 DNA carrying cloned *pdc* and *adhB* genes from *Z. mobilis* and a bacterial gene allowing tetracycline resistance.
- Digest the isolated DNA with restriction endonucleases.
- Analyze the restriction enzyme digested and undigested pLOI297 using agarose gel electrophoresis.
- Map the position of the restriction endonuclease sites on pLOI297.
- Analyze the DNA sequence of pLOI297 using bioinformatics tools available publicly online.

### Introduction

Our ability to clone, sequence and analyze genes from any organismal source has provided us with a wealth of information about the biology of organisms and their relationship to each other. The cloning of genes occurs in a two-step process: 1) recombinant DNA is produced by attaching the DNA of interest to DNA from another source, often DNA that has the properties of a chromosome; and 2) the recombinant DNA is cloned through introduction into a cell that recognizes the presence of the chromosomal signals and replicates the DNA into many identical copies (the “clones”). Our ability to recombine and clone DNA is due to advances made in understanding bacterial genetics and the enzymatic reactions of DNA. (See pages 518-524 and 628-635 in your textbook.)

Scientists immediately recognized that gene cloning would allow genetic engineering – using the cloned genes to solve practical problems facing the world. A cloned gene could be introduced into an organism to provide it with new properties, for example, the introduction of the gene for the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) into crop plants to make them resistant to the herbicide Roundup. The protein product of a cloned gene could be produced for medical use, for example the production of human insulin to treat diabetes. (See pages 635-640 in your textbook.) Developing sources of renewable liquid energy to limit our dependence on fossil fuels is one of the practical problems facing the world today. Although annual global production of biofuels is now over 100 billion liters a year, it represents only 2% of the fuel used in global transportation. One way scientists are addressing this problem is by developing biofuels (bioethanol or biodiesel) through the use of genetic engineering.

This lab will be done over two periods in which you will learn several common techniques used by molecular geneticists in any study of gene structure and function. You will perform these techniques on pLOI297, a plasmid that was generated to engineer *E. coli* cells to overproduce ethanol through the process of fermentation. pLOI297 was the product of recombinant DNA reactions with pUC18 to include the *Zymomonas mobilis* genes for pyruvate

---

<sup>1</sup> The LI-COR sequencer was awarded to Gustavus by LI-COR, Inc.

decarboxylase (*pdc*) and alcohol dehydrogenase (*adhB*), and a bacterial gene from the *E. coli* plasmid pBR322 that allows resistance to the antibiotic tetracycline.

## Common Techniques in Recombinant DNA Work

**Vectors and cells.** Using restriction endonucleases and DNA ligase, molecular biologists are able to insert pieces of DNA into **vectors** such as plasmids or viruses. The plasmid pUC18 (see map below) is specially designed for easy gene cloning. Like many plasmids used as cloning vectors it 1) is “relaxed” so that many copies of the plasmid can be found in one bacterial cell; 2) possesses distinguishable genetic markers that allow selection of cells containing the plasmid and of cells containing plasmid with inserted DNA; 3) has several cloning sites, i.e. sites where restriction enzymes cut only once; and 4) has a low molecular weight. Size is important since small plasmids can contain larger gene inserts and still be stable molecules inside the bacterial cells.

Special mutant strains of *E. coli* are designed for work with recombinant DNA molecules. For example, strain TC4 containing pLOI297 is unable to grow unless provided with a complete medium and therefore cannot grow if it “escapes” the lab. It also contains a deletion in the  $\beta$ -galactosidase gene of the *lac* operon which destroys its ability to utilize lactose as a carbon source and allows it to participate in a color screen to identify bacterial colonies that contain recombinant molecules.

**Selection.** In order to clone the inserted DNA, the vector with its inserted DNA must be introduced into cells, and the cells carrying the introduced DNA must be identified. The identification of cells containing newly introduced DNA is called **selection**. The type of selection used varies depending on the vector and cell type used in the experiment. In this lab we use plasmid vectors introduced into *E. coli* cells, the most common combination used by molecular biologists.

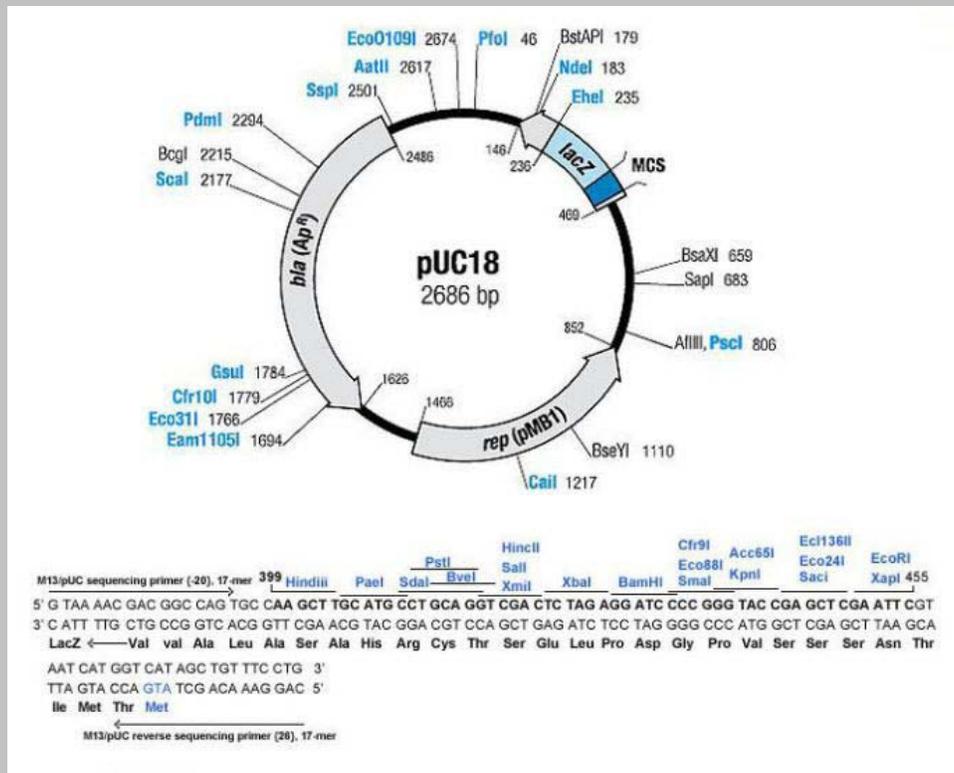
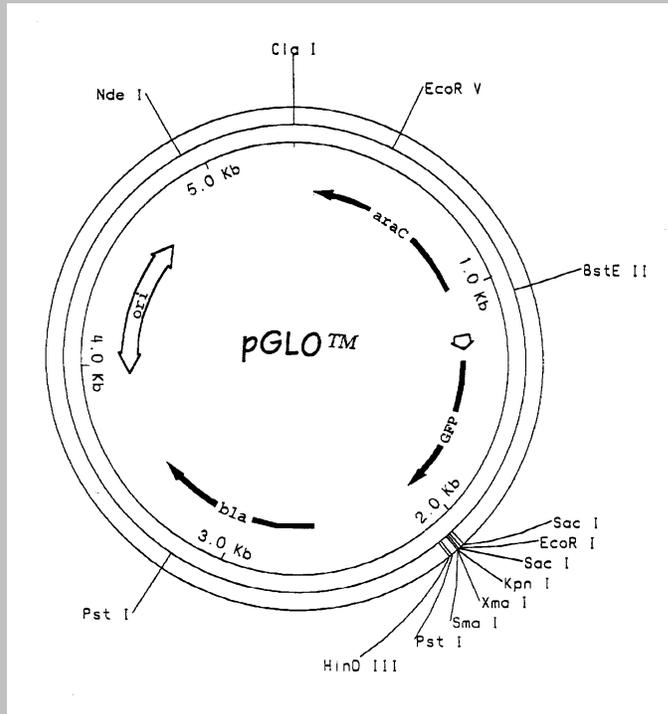
**DNA isolation.** The procedure used to isolate DNA varies from organism to organism and cell type to cell type. Different methods of disrupting cells and tissues are necessary when working with organisms as disparate as plants, mammals, and bacteria. For recombinant DNA experiments, DNA must be isolated away from contaminating carbohydrates, nucleases that would degrade the DNA, and other cellular molecules that may inhibit the specific enzymes used in recombinant DNA experiments. Plasmid DNA is isolated from bacterial cells by lysis of the cells. The bacteria's chromosomal DNA and proteins are selectively removed by precipitation and the remaining cleared lysate contains the plasmid DNA as well as cellular RNA. The plasmid DNA is concentrated by precipitation in alcohol.

**Restriction endonuclease digestion.** As described on pages 518 and 520-521 (Box 18B) of your textbook, restriction endonucleases or enzymes digest or “cut” DNA at specific sequences generating either blunt or “sticky” (i.e. single-stranded) ends. Digestion of DNA with restriction endonucleases results in information about the structure and sequence of DNA; restriction endonuclease sites act as the “cities” on a “highway” of mapped DNA. Restriction endonuclease sites are used to provide compatible sites to introduce fragments of DNA into a vector, but have also been used as genetic markers for inherited diseases and in “fingerprinting” crime suspects. The position of some restriction endonuclease sites are shown on the pUC18 and pGLO maps below.

**Gel electrophoresis.** Gel electrophoresis is used to separate DNA fragments by size (p. 519, Fig. 18-12 of your textbook). Fragments of a wide range of sizes, for example over several kilobases in length, are separated on agarose gels. By comparing the distance of migration of unknown DNA fragments with the migration distance of marker fragments of known size, the size of a gene and the position of the recognition sites for restriction endonucleases can be determined. Acrylamide gels are used to separate fragments different in size by one or only a few base pairs, for example, the products of DNA sequencing reactions described below.

**DNA sequencing.** The most commonly used method for sequencing DNA is the chain termination or Sanger technique, named for Fred Sanger, the biochemist who won the Nobel prize for development of the method. This

technique relies on the requirements for DNA replication – a primer, deoxynucleotides with a 3' OH group and a DNA polymerase that will synthesize DNA by adding a new nucleotide onto the 3' OH group of the growing chain.



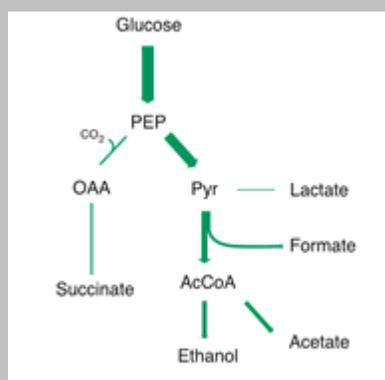
[http://www.genscript.com/vector/SD1162-pUC18\\_plasmid\\_DNA.html](http://www.genscript.com/vector/SD1162-pUC18_plasmid_DNA.html)

## Biofuels and pLOI297

The energy rich molecules currently used as liquid biofuels are biodiesel (lipid-based) and ethanol (alcohol-based). The most cost-efficient way to obtain biofuels is through the conversion of biomass. Biodiesel is mainly produced using triacylglycerol obtained from crop plants that are high oil producers (e.g. soy, canola and palm) whereas ethanol (or bioethanol) is produced in large part from the fermentation of corn and sugarcane. With current production processes, ethanol is a more efficient choice as it requires less land per unit of biofuel produced than biodiesel.

The use of crop plants for biofuels is problematic. The fuel market competes directly with the food market for these resources, driving up prices on both sides. Increased demand for crop plants drives an increased need for arable land, resulting in increased deforestation. In order for biofuels to offer a long-term solution to our energy needs, scientists are investigating numerous ways to produce biofuels. Instead of using the corn seeds produced by corn plants for ethanol production, scientists are working to improve methods for purifying cellulose from the remaining plant material after harvest. Plant cell walls are largely made up of cellulose, hemicellulose and lignin; the cellulose provides the sugars necessary for fermentation. So-called “cellulosic” ethanol production could utilize a variety of non-food crop plants too, reducing the need to expand cultivatable acreage and eliminating the “food or fuel” conundrum.

Another active area of research involves developing microorganisms such as algae and bacteria for production of biofuels. Professor Lonnie Ingram of the Florida Center for Renewable Chemicals and Fuels at the University of Florida has approached the fuel problem by investigating the ability of microorganisms to produce bioethanol, specifically engineering *E. coli* to produce high levels of ethanol. Bacteria that are able to live under anaerobic conditions must have the ability to produce ATP through the fermentation process. *E. coli* is a facultative aerobe (able to grow under aerobic or anaerobic conditions) and is the bacterial host of choice for performing gene cloning and obtaining the protein expressed from the cloned gene(s). But *E. coli* utilizes a fermentation pathway known as “mixed-acid fermentation” that results in the production of acids such as lactate, acetate and succinate in addition to the ethanol that is the major product of yeast fermentation (see diagram below) and thus is not naturally a suitable organism for ethanol production.



<http://www.bio.anl.gov/people/microbiology/microbiology2.html>

To generate *E. coli* cells that would produce ethanol at the high and efficient levels necessary, Dr. Ingram and his collaborators cloned the two genes necessary for ethanol production from pyruvate – the genes for pyruvate decarboxylase and alcohol dehydrogenase, two enzymes that should be familiar to you from our unit on metabolism – from *Z. mobilis*, a bacteria that follows alcoholic fermentation. The Ingram lab hypothesized that if these cloned genes were expressed by the *E. coli* cells at a high level, the *E. coli* would become mainly ethanol producers at the expense of the acids they had been producing. The plasmid pLOI297 was a product of this work and has been shown to result in efficient ethanol production when the *E. coli* containing the plasmid are grown on glucose, lactose or xylose (Alterthum and Ingram, 1989).

(Note: The name pLOI297 represents Dr. Ingram's initials followed by a number assigned by the Ingram lab during the plasmid's development. Many recombinant DNAs are named in this fashion, for example, pUC18 is named for the University of California where it was first generated.)

## References

Altherthum, F. and Ingram, L.O. (1989) Efficient Ethanol Production from Glucose, Lactose, and Xylose by Recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* 55:1943-1948.

Youngs, H. and Somerville, C. (2012) Growing Better Biofuel Crops. *The Scientist* July 2012, pp. 46-52. Retrieved from <http://the-scientist.com/2012/07/01/growing-better-biofuel-crops/>

## **Week 1: Selection, DNA Isolation and Restriction Endonuclease Digestions**

### **Isolation and Restriction Endonuclease Digestion of pLOI297**

In the lab today, you will isolate pLOI297 from *E. coli* TC4 cells and digest your DNA with three different restriction endonucleases. pLOI297 DNA was created in a multistep process by digesting pUC18 DNA with a restriction enzyme and inserting the three genes (*pdv*, *adhB*, and *tet<sup>r</sup>*) sequentially into the open plasmid using DNA ligase. The three cloned genes are between the *EcoRI* and *PstI* sites shown on the pUC18 map above. Once you have isolated the pLOI297 DNA, you will digest it with the enzymes *EcoRI*, *PstI* and *Sall*, singly and in combination. Next week you will electrophorese the DNA and determine where those enzymes digest the cloned fragment of DNA and determine a restriction map of the cloned gene.

*\*Why does pLOI297 meet the definition of "recombinant DNA"?*

1. Obtain a tube containing an overnight culture of *E. coli* TC4 cells carrying pLOI297. Add 1.5 mL of the culture to each of two microcentrifuge tubes. Cap the tubes and label them with a Sharpie or other permanent marker. Centrifuge the tubes for one minute in the microcentrifuge (15,900xg) to pellet the cells.
2. Pour off the broth from each tube and drain the tubes on a paper towel. If necessary, use a micropipet and sterile tip to remove the last drops of broth from the tube. **Save the bacterial pellet.**
3. Resuspend each pellet in 0.2 mL GTE buffer (50 mM glucose, 25 mM Tris pH 7.5, 10 mM EDTA) containing 1 mg/mL lysozyme. The GTE buffer is isotonic to the cells; the lysozyme will digest the bacterial cell wall. **Make sure the pellet is disrupted completely into a suspension of single cells by finger flicking the tube or pipeting the solution up and down.** Keep the tube at room temperature for 5 minutes.
4. Add 0.4 mL of NaOH:SDS solution (0.2 M NaOH, 1% SDS) to each tube. Mix gently but thoroughly by inverting the tube several times. The solution should become viscous. Place the tube on ice for 5 minutes. The SDS will lyse the cells and the NaOH will denature DNA and hydrolyze RNA.
5. Add 0.3 mL of ice-cold 7.5 M ammonium acetate to each tube. Mix gently by inverting the tubes several times. Incubate the tubes on ice for 10 minutes. The ammonium acetate solution will neutralize the alkaline NaOH and precipitate the SDS, forming a chunky white precipitate. Chromosomal DNA, cellular proteins and other cellular debris are found in the precipitate, whereas the plasmid DNA and cellular RNA are not.
6. Centrifuge the tubes for 5 minutes in the microcentrifuge. **The supernatant contains the plasmid.** Transfer the supernatant to two new microcentrifuge tubes using a micropipet and sterile tip without disturbing the pellet or transferring any cellular debris. The original tubes containing the precipitate can now be discarded.

7. Add 5  $\mu$ l of RNase A (100 $\mu$ g/mL) solution to each tube. Mix thoroughly and incubate the tubes in a 37°C water bath for 5 minutes. RNase A will digest the remaining cellular RNA.
8. Add 0.6 volume of isopropanol to each tube relative to the volume of recovered supernatant (approximately 0.4 - 0.5 mL of isopropanol). Mix thoroughly (gentleness no longer required) and leave at room temperature for 10 minutes. The solution may become cloudy due to precipitation of the nucleic acids in the presence of monovalent cations.
9. Centrifuge the tubes at **room temperature** for 10 minutes to pellet the precipitated plasmid DNA. Place the tube **hingside up** in the microcentrifuge so that the DNA pellet can be detected. The plasmid is in the small white pellet at the bottom of the tube or trailing up the side of the tube. Decant the isopropanol and drain the tubes on a paper towel.
10. Wash the DNA pellets each with 300  $\mu$ l 70% ethanol and finger-flicking to remove excess salt from the pellet. Centrifuge the tubes for 5 minutes. Pour off the 70% ethanol and drain the tubes on a paper towel. Remove any remaining drops of ethanol using a micropipet and sterile tip. Be careful not to disturb your DNA!
11. Allow the pellets to air dry for 15 minutes. The pellet can be damp, but the alcohol should be evaporated before continuing. Once the pellets have dried, resuspend the DNA in 50  $\mu$ l TE (10 mM Tris, pH 8.0, 0.1 mM EDTA) each by vortexing. Allow resuspension to occur for at least 5 minutes before going on to the next step. Incubating at 37°C may help resuspension.

12. Obtain six microcentrifuge tubes and label them with permanent marker for the six restriction enzyme reactions you will perform:

1) *EcoRI*    2) *PstI*    3) *SalI*    4) *EcoRI* + *PstI*    5) *EcoRI* + *SalI*    6) *PstI* + *SalI*

Include your initials or some other identifying mark so that your group will be able to distinguish your tubes from all the other tubes in the class.

13. Using a micropipet and sterile tip, pipet 5  $\mu$ l of DNA from one of your tubes into each of the six marked tubes. **DO NOT THROW AWAY YOUR TUBES OF UNUSED DNA!** Give your two microcentrifuge tubes of remaining DNA to your lab instructor and he/she will store the tubes of DNA in the freezer until the next lab period.
14. Six tubes, each containing a mixture of the reaction buffer and the restriction enzyme(s) for a particular reaction, will be in an ice bucket in the lab. **Leaving the buffer + enzyme tubes on ice while you work with them,** pipet 15  $\mu$ l of the corresponding mix into your labeled DNA-containing tube. The enzymes will lose their activity with time if not kept on ice. **Change the micropipet tip between each buffer + enzyme tube!** Make sure the buffer + enzyme and DNA are mixed by finger-flicking your microcentrifuge tube; collect the reaction at the bottom of the tube by spinning for 1-2 seconds in a microcentrifuge. The final reaction conditions in your tubes will be as follows:

- pLOI297
- 10 units of restriction enzyme
- 1x Restriction Endonuclease Buffer (Buffer H: *EcoRI*, *PstI*, *EcoRI* + *PstI*; Buffer D: *SalI*, *EcoRI* + *SalI*, *PstI* + *SalI*)

The composition of these buffers is shown in the table below:

Buffer	pH (at 37° C )	Tris-HCl mM	MgCl <sub>2</sub> mM	NaCl mM	KCl mM	DTT mM
A	7.5	6	6	6	-	1
B	7.5	6	6	50	-	1
C	7.9	10	10	50	-	1
D	7.9	6	6	150	-	1
E	7.5	6	6	100	-	1
F	8.5	10	10	100	-	1
G	8.2	50	5	-	-	-
H	7.5	90	10	50	-	-

\* Why might different restriction enzymes require different salt conditions?

15. Incubate the tubes in the 37°C water bath for the remaining time in lab, 1.5 - 2 hours. Your lab instructor will remove the tubes at the appropriate time if the incubation runs past the end of lab.

### Selection of cells containing cloned DNA

Selection of *E. coli* cells containing plasmids is performed in several ways. One way is antibiotic resistance. Both pUC18 and pGLO carry a gene for ampicillin resistance (see maps above). *E. coli* cells are normally ampicillin sensitive, therefore cells containing these plasmids can be selected by growing on ampicillin-containing media. The pUC18 vector also carries part of the *lacZ* gene (see map). The *lacZ* gene encodes the β-galactosidase necessary for lactose utilization by the cell. *E. coli* cells are normally *lacZ*<sup>+</sup>, but TC4 and DH10B cells are *lacZ*<sup>-</sup> and cells containing pUC18 can be selected by color on special indicator plates. MacConkey plates differentiate between bacterial cells that can utilize lactose (deep pink to purple color) and those that cannot utilize lactose (colorless or very pale pink and translucent) due to the presence of Neutral Red dye as a color indicator. As fermentation of lactose occurs in *lac*<sup>+</sup> cells, lactate and other acids are produced and the pH drops thus producing the deep color. This does not occur in *lac*<sup>-</sup> cells.

Another type of selection is possible if the *E. coli* cells produce the protein encoded by the cloned gene. Colonies producing the protein can be identified and isolated away from other colonies not producing the protein. In these cases the gene encoding the protein is called a “reporter gene” – the presence of the protein “reports” that the cloned gene is present in those cells. Cells containing pGLO produce green fluorescent protein (GFP) and the fluorescence of the colony “reports” to you that the pGLO DNA is present in those cells.

As can be seen on the pUC18 map, the *EcoRI* site is within the *lacZ* gene. When a gene is inserted at this point, the *lacZ* gene becomes inactivated through **insertional mutagenesis**, and TC4 cells carrying this plasmid can no longer utilize lactose as a sugar source. You will perform a color selection by plating TC4 cells with pLOI297 and DH10B cells with pUC18 on MacConkey agar. The agar will contain different antibiotics for selection by antibiotic resistance. In addition, you will plate out a mixture of cells, some of which contain pGLO, to demonstrate the use of GFP as a reporter gene.

1. Obtain from your lab instructor the following dilute bacterial cultures:
  - 1) DH10B
  - 2) DH10B carrying pUC18 (“pUC18”)
  - 3) TC4 carrying pLOI297 (“pLOI297”)
  - 4) Mixture of cells carrying pUC18 and cells carrying pLOI297
  - 5) Mixture of cells carrying pUC18 and cells carrying pGLO

2. Obtain 4 of each of the following MacConkey agar plates:

- 1) MacConkey
- 2) MacConkey + ampicillin
- 3) MacConkey + tetracycline
- 4) MacConkey + kanamycin

Obtain 1 plate of LB (Luria-Bertani) agar containing ampicillin + arabinose

3. Plate 0.1 mL of bacterial cultures 1 to 4 on one of each type of MacConkey plate. To plate the cells, use the "hockey stick" method - dip the flat end of the glass spreader in alcohol, pass the spreader through a Bunsen Burner, and allow the alcohol to burn off. Use the sterile surface of the spreader to spread the cells in a thin film across the surface of the plate. Pipet and spread one plate at a time – if you pipet onto too many plates in advance the cells will stick and you will get a “blob” instead of distinct colonies. Also plate a 0.1 mL sample of culture 5 on the LB (amp + ara) plate. Make sure to label the *agar* side of the plate with your name and lab section.

\* What would you predict for the color of DH10B cells containing pUC18 carrying a cloned gene? pUC18 without a cloned gene? Untransformed DH10B cells?

\* Which cells will grow on ampicillin? Kanamycin or tetracycline? Plates without antibiotics?

\* Which plates will contain fluorescent cells under ultraviolet (UV) light?

4. Grow the plates up overnight in the 37°C incubator (*agar side up*). It is important not to leave the plates for more than 18 hours. The plates can be stored in the refrigerator until the next lab period.

5. Observe the colonies on each plate and determine their color. The red-purple color should be quite dark; the cream colored colonies may look pastel pink or violet due to the color of the MacConkey plate. Observe the colonies on the LB (amp + ara) plate under UV light and note the ability of the colonies to fluoresce. The fluorescence pattern will either be positive, negative, or a mixture of both colony types. **Do not look directly at the UV light source!**

**Type of colonies (record color and growth/no growth):** Complete the “pUC18 + pGLO” column with your predictions for growth and color on the plates indicated.

	DH10B	“pUC18”	“pLOI297”	“pUC18” + “pLOI297”	“pUC18” + “pGLO”
Plate type:					
<b>MacC +Ampicillin</b>					
<b>MacC + Kanamycin</b>					
<b>MacC +Tetracycline</b>					
<b>No antibiotics</b>					
<b>LB + Amp + Arabinose</b>					
<b>Fluorescence Pattern</b>					

## Week 2: Agarose Gel Electrophoresis of DNA and DNA Sequence Analysis

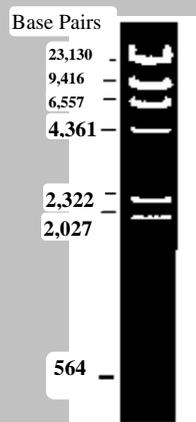
Today you will use gel electrophoresis to separate the DNA fragments that were produced in the restriction endonuclease reactions of your pLOI297 DNA. You will determine the sizes of the fragments and use this data to produce a map of the enzyme sites in the cloned DNA. You will also perform DNA sequence analysis on pLOI297 sequence obtained with the LI-COR sequencer and/or another sequencing source, using these results to determine the order of the genes inserted in pUC18.

### Electrophoresis of digested and undigested pLOI297 DNA

1. Wash a Lucite gel bed and well former or "comb" with deionized water. Following the directions for the electrophoresis unit (See "3.1 Mini-Gel Unit Preparation and Gel Casting" at the end of the lab), set up the unit to cast, or pour, your gel.
2. Obtain a flask with 40 mL of 1 % agarose gel solution containing 1x TAE (50 mM Tris-acetate (pH 7.5), 1 mM EDTA), an electrophoresis buffer, from the 70°C water bath. Add 5 µl of the fluorescent dye ethidium bromide (10 mg/mL). **Ethidium bromide is a carcinogen.** The student(s) working with the gel must wear disposable gloves. The ethidium bromide will bind to the DNA in the gel as the DNA is electrophoresed.

Immediately after adding the ethidium bromide, swirl the flask several times and pour the gel solution into the gel bed as described. If large bubbles are present they can be popped with a micropipet tip.

3. Defrost your digested and unused DNAs from the previous lab. Add 2 µl of the 10x loading buffer (0.25% Bromophenol Blue, 0.25% Xylene Cyanol, 30% glycerol in TAE) to each tube containing **digested** DNA. Pipet 5 µl of your stock, **undigested** DNA, 13 µl 1x TAE buffer, and 2 µl of 10x LB into a new microcentrifuge tube. The loading buffer contains two dyes, bromophenol blue and xylene cyanol that will allow you to monitor the progress of the electrophoresis. Bromophenol blue will migrate at approximately the same position as tRNA molecules.
4. Once the gel has solidified (approx. 15-30 minutes), carefully pull the well former out of the gel as described in "3.2 Running the gel." The wells should be closest to the black lead connection. Add 1x TAE buffer to the box so that the gel is covered by 1-2 mm of buffer. Load approximately 20 µl of each of your samples in wells side by side on the gel using a micropipettor. Record the order in which the samples were loaded onto the gel. Further loading "hints" will be given in lab. Your lab instructor will provide you with lambda DNA (0.1 mg/mL) digested with *Hind*III as size standards for the gel - load 20 µl of the prepared sample. Eight DNA fragments of known size are present in the lambda DNA standard provided; you will probably see 7 of the 8 fragments. The gel pattern and sizes of the fragments are shown below.

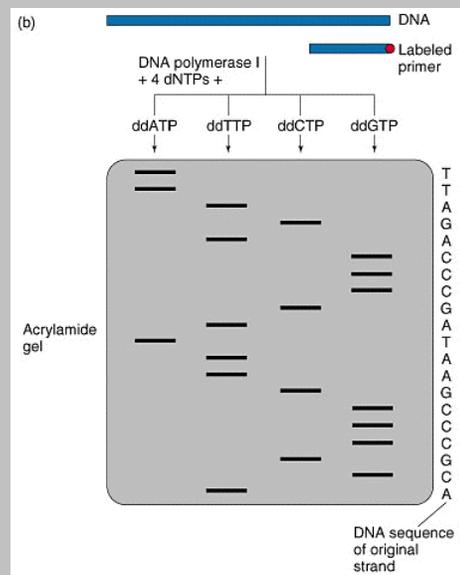
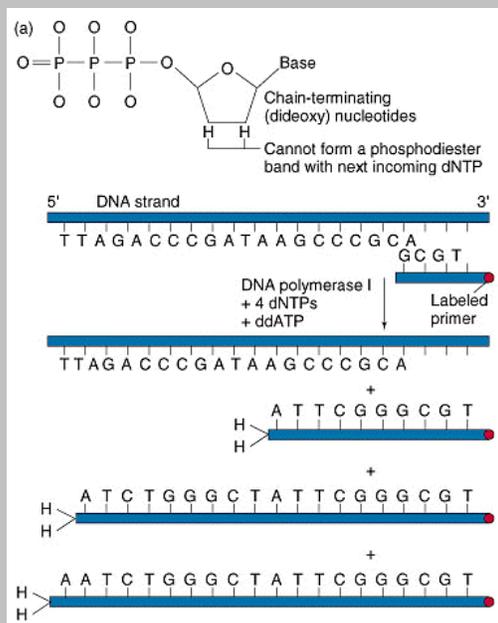


Sizes of Lambda DNA fragments from digestion with *Hind*III are given in base pairs.

- Once the samples are loaded, the gel will be electrophoresed at 120 – 150 volts (which should be no more than 40 -50 milliAmps for one gel; two gels on the same power source can be at 80 -100 milliAmps). Directions for the use of the power supplies were included with the protein electrophoresis lab. Remember that DNA is negatively charged and will move towards the positive anode (red colored lead). Typically, electrophoresis continues until the bromophenol blue front nears the end of the gel, about 40 – 60 minutes.
- Once electrophoresis is complete, remove the gel from the electrophoresis unit and place it on the UV transilluminator. Cover the transilluminator with the plexiglass shield before turning on the UV light! The presence of DNA is detected by the fluorescence of bound ethidium bromide. If the gel looks orange throughout, the gel needs to be destained in H<sub>2</sub>O. If the 564 bp lambda DNA fragment cannot be seen, your gel may need additional staining in order to see all the fragments in your experimental lanes.
- Place a ruler next to your gel and photograph the gel using the gel documentation system. Use the Image J software on the computer to capture an image (directions are by the computer). Copy the image onto a flash drive or email it to yourself for further analysis.
- Dispose the gels and electrophoresis buffer in containers found in the fume hood.

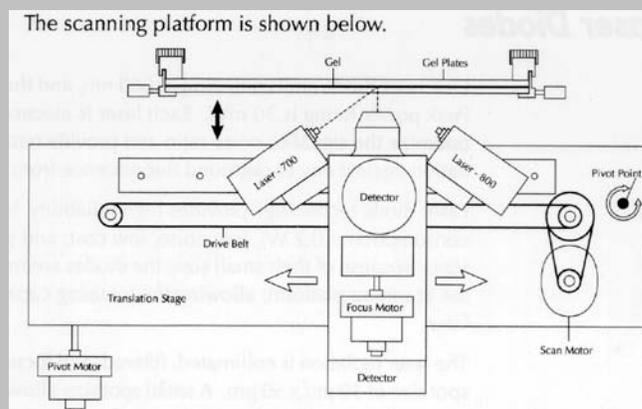
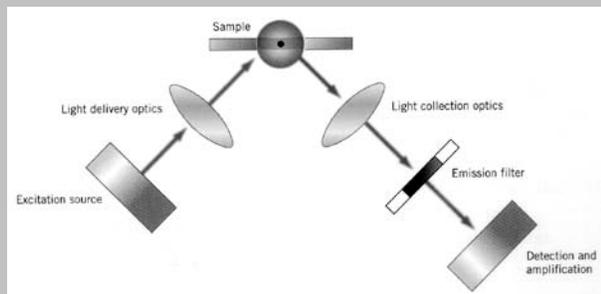
## DNA Sequence Analysis of pLOI297

The human genome was sequenced using the chain termination or Sanger technique, named for Fred Sanger, the biochemist who won the Nobel prize for development of the method. This technique relies on the requirements for DNA replication – a primer, deoxynucleotides with a 3' OH group and a DNA polymerase that will synthesize DNA by adding a new nucleotide onto the 3' OH group of the growing chain. As shown in the diagram below left, a primer with a label (radioactive or fluorescent) on its 5' end base pairs to the DNA region to be sequenced. All four normal deoxynucleotides and a DNA polymerase are also present. This combination is split into four separate reactions, each with one of the four dideoxynucleotides, a deoxynucleotide without the 3' OH group. As in the example reaction shown below, each time an A needs to be added to the DNA chain, the polymerase has a choice between the dATP or the ddATP. If the ddATP is added, the chain cannot be lengthened and stops with an “A” base at the final position.



All four reactions for one starting DNA are electrophoresed on an acrylamide gel that separates DNA fragments with single nucleotide differences between fragment lengths (diagram above right). The label on the primer allows visualization of each fragment with one of the four didoxynucleotides on the end as a band. The sequence can be read from the bottom of the gel towards the top – using a straight edge you move from the bottom band (“T” in the example) to the lane where the next highest band is in the gel (“G”) and continue this process as far up the gel as possible (for example, “C” is next). The original strand of DNA had the *opposite* sequence due to the specificity of A:T and G:C base pairing.

Automated sequencers “read” the position of fragments in the gel continuously and are interfaced with a computer whose software captures, saves and analyzes the data. The LI-COR automated sequencer detects primers in the sequencing reactions labeled on the 5' end with an infrared dye. The reactions are electrophoresed on a vertical gel and as electrophoresis occurs, the DNA fragments separate by size and pass by a “window” near the bottom of the gel. Behind this window is a laser diode similar to that found in a CD player. The laser excites the dye attached to any DNA fragment appearing in the window at that moment and a focused fluorescent microscope that moves back and forth across the window detects the fluorescence of the dye. As is the case for most fluorescent dyes, the excitation wavelength is different from the wavelength that produces maximal fluorescence for detection. For example one dye used in this system, IRD700, absorbs maximally at 685 nm and emits fluorescence maximally at 710 nm. The diagrams below show the basis of fluorescence imaging (left) and the apparatus behind the scanning window (right).



Once the DNA sequence is determined, it is often compared to all the DNA sequences deposited in the public databases. A common algorithm used for this analysis is BLAST (Basic Local Alignment Tool). The direct comparison produced by BLAST may allow the identification of the gene if it closely matches the sequence of another known gene. The differences between the same genes from different organisms - for example, hexokinase, the first enzyme in glycolysis or the rRNA genes – can help establish the evolutionary relatedness of different species. The comparison of a gene sequence from a patient with the sequence from a normal individual may identify a specific disease-causing mutation.

1. Prior to this lab, DNA was isolated from a culture of cells identical to yours, and the DNA sequencing reactions and electrophoresis was performed on the recombinant pLOI297. Your lab instructor will take you to the biochemistry lab (NHS 207) to view the sequencer and the sequencing results if scheduling permits. You will be given .txt files of the sequence for your DNA as determined by the LI-COR sequencer and its associated software, e-Seq. You will also be told where you can access the graphic results of sequencing in the form of a gel “picture.” - Sequences will be from “the 700 channel” or “the 800 channel.” Sequences from one channel are a reading of the sequence from one direction (one strand) and sequences from the other channel are a reading from the other direction (the complementary strand).

2. Enter the sequences of your DNA into a Word file. The standard font used for DNA sequence analysis is “Courier New.” Save this file for future reference.
3. Find the four lanes that represent your DNA on the “gel picture.” Using a ruler or straight edge, read approximately 100 nucleotides of the sequence from the bottom towards the top of the picture.
4. Find the sequence you read from the picture within the sequence you entered into your Word file. How many differences are there between your reading of the sequence and that determined by the e-Seq software? Are these differences due to mistakes by you, e-Seq, or both? **Indicate the overlapped sequence and points of differences in a figure for your lab report.**
5. Go to The National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Click on “BLAST” in the list of “Popular Resources” on the right-hand side of the NCBI home page. The following page should come up:

**BLAST** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** Aligning Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool. [Go](#)

**BLAST Assembled RefSeq Genomes**

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

**Basic BLAST**

Choose a BLAST program to run.

<a href="#">nucleotide blast</a>	Search a <b>nucleotide</b> database using a <b>nucleotide</b> query <i>Algorithms:</i> blastn, megablast, discontinuous megablast
<a href="#">protein blast</a>	Search <b>protein</b> database using a <b>protein</b> query <i>Algorithms:</i> blastp, psi-blast, phi-blast

**News**

[BLAST 2.2.23 release](#)

A new version of the stand-alone applications is available.  
Mon, 22 Mar 2010 15:00:00 EST

[More BLAST news...](#)

**Tip of the Day**

[More tips...](#)

6. The BLAST search will compare your DNA sequence to that of all sequences currently within GenBank. There are many different types of BLAST searches that can be done, but we are going to begin by doing a basic BLAST search of nucleotides sequences. To do this, click on “nucleotide blast.” The following page should come up:

Enter Query Sequence [Reset page](#) [Bookmark](#)

---

Enter accession number, gi, or FASTA sequence [Clear](#) Query subrange From   
To

Or, upload file  [Browse...](#)

Job Title   
Enter a descriptive title for your BLAST search

**Choose Search Set**

Database  Human genomic + transcript  Mouse genomic + transcript  Others (nr etc.):  
Human genomic plus transcript

Entrez Query   
Optional  
Enter an Entrez query to limit search

**Program Selection**

Optimize for  Highly similar sequences (megablast)  
 More dissimilar sequences (discontiguous megablast)  
 Somewhat similar sequences (blastn)  
Choose a BLAST algorithm

**BLAST** Search database **Test/gpipe/9606/allcontig\_and\_rna** using **Megablast (Optimize for highly similar sequences)**  
 Show results in a new window

[Algorithm parameters](#)

7. On this page, we can enter in our unknown DNA sequence. Copy and paste your sequence into the box under the “Enter Query Sequence” tab. We will now do a broad search in order to see which nucleotide sequences most closely match our unknown DNA sequence. To do this, under the “Choose Search Set” tab, select “Other (nr etc.)”. Under the “Program Selection” tab, select “Somewhat similar sequences (blastn)”. Finally, click on “BLAST.” - You will need to perform the BLAST comparison three times, once for each of three sequences: 1) your 100-nucleotide manually read sequence from either the 700 channel or the 800 channel, 2) your machine-read sequence from the 700 channel, and 3) your machine-read sequence from the 800 channel.

8. The computer will return an analysis reporting the sequences that come closest to matching your sequence. They will be reported in descending order from best to worst match within certain parameters established by the program. The E value is especially important – this is the probability of alignments with equivalent or superior score that could have occurred purely by chance. E scores of 0.01 or lower are considered statistically significant. A partial name and active links for each sequence are present in the list. Click on the accession number to see the complete description for a particular sequence.

Scroll down past the list to see the nucleotide by nucleotide match-up between your sequence and those sequences reported in the list. Some reported sequences will be the individual gene, some may be the entire genome where the sequence is found, and still others may be cDNAs or ESTs. Alignments with lower E values will involve more matching nucleotides – alignments with higher E values (e.g. 0.01) will involve smaller numbers of nucleotides, as would be expected if there was a higher probability of a match by chance.

9. By clicking on the active links to specific sequences, determine what protein each sequence encodes. Most of the specific sequences you will retrieve will be “annotated,” meaning that someone has analyzed the sequence for its probable product, the start and stop codons, etc. Record an “accession number” of a complete gene for each of you sequences to use in step 12 below.

10. By looking at your rank ordered list of matches and the nucleotide by nucleotide matches, determine which organisms have a similar gene within their genomes. Do you obtain only bacterial sequences or do plants and animals also have a gene encoding the same or similar protein? Indicate the approximate similarity between the sequences you obtain by indicating the extent of the nucleotide match (% and number of bases) for a representative selection of organisms.

*\*Would you expect other plants, animals, bacteria, etc. to have similar sequences? Why or why not?*

11. This was the most general search that we could do. Now let’s try to do some more specific searches. By using the “Back” icon, return to the screen shown on the previous page. Now, under the “Program Selection” tab, select “Highly similar sequences (megablast).” Click on “BLAST.” How do you think that these results will differ from the previous results?

**Note:** If you want to save any of your nucleotide by nucleotide alignments by copying and pasting into a Word document, make sure to use Courier New as the font so that the alignment is maintained.

12. Your DNA sequence analysis can be tied with your restriction enzyme analysis by locating the restriction enzyme sites found in the individual gene sequences. Go to the New England Biolabs NEBcutter program (<http://tools.neb.com/NEBcutter2/index.php>) and input the accession number for one of your sequences in the “GenBank Number” box. Click on “Submit” – NEBcutter will return a map of the DNA indicating the position of all restriction enzyme sites. Under “Main Options” select “Custom digest” and choose specific enzymes with which to digest your DNA sequence. Are these results consistent with what you found from the gel electrophoresis of your digested DNA?

# Data Analysis

1. **Selection.** Using the growth and colony color information obtained from selection of cells on MacConkey agar, describe the genotype of the different cells relative to antibiotic resistance and lactose utilization. Did your results match your predictions? What predictions can you make for the growth and color of colonies if the “pUC18 + pGLO” mixture was plated on the four types of MacConkey agar?
2. **DNA isolation and restriction enzyme digestion.**

- a. Compare the gel lane containing your undigested DNA with the lanes containing your restriction endonuclease digestions. When plasmid DNA is isolated from bacterial cells, several physical forms of the same DNA molecule are found: supercoiled circles, nicked circles, linear, and multimeric forms. Molecules containing the same number of base pairs but in different physical forms migrate differently in gel electrophoresis. When these molecules are digested to completion with the restriction endonucleases, only discrete linear forms are found.

*\* How does the physical nature (circular v linear) of a DNA molecule or fragment affect the migration distance of the molecule through the gel?*

- b. Use Sigma Plot to make a semi-logarithmic plot of the distance of migration (mm) of the linear lambda DNA fragments (linear scale; x axis) v the size of the molecules in base pairs (log scale; y axis). Depending on the concentration of agarose in the gel, there is a semilog relationship between the distance migrated and the size of the DNA. For a 1% agarose gel, linear DNA fragments between 0.5 and 7 kb will migrate in a linear fashion. Use this curve to determine the sizes of the linear DNA fragments from your six different restriction enzyme digests.

Using the **known** information about pUC18 (see map) and the information from your experiment, determine the size of the pLOI297 DNA insert (i.e. the length of the fragment that contains *pd*c, *adhB* and *tet<sup>r</sup>*), and the presence and position of sites for the three restriction enzymes. Generating a restriction enzyme map is a lot like piecing together a puzzle. Remember that your pUC18 and cloned genes do not change in size with different restriction enzymes, i.e. the added length of all the fragments in the *Sa*II reaction should be the same or similar to the added length of all the fragments in the *Eco*RI reaction.

3. **DNA sequencing.**
  - a. Based on the BLAST results of your sequences, what gene is nearest the *Eco*RI site of pUC18? The *Pst*I site? What other organisms have similar genes? How much similarity for these genes exists between species?
  - b. Combine your results from both your restriction mapping and analysis of pLOI297 to determine the order of the three genes inserted into pUC18 relative to the *Eco*RI and *Pst*I sites.

## Lab Report

**For this laboratory you are expected to turn in a titled report containing an introduction, materials and methods, results and literature cited section. Your report should be written in the style of a molecular biology journal article using the guidelines below. Your audience is someone who is in or has already completed BIO 201 or has the equivalent or greater molecular biology experience.**

This assignment is worth 20 points and will be due when your lab section meets for Lab 11. Although you may perform the data analysis with your lab partner(s), **the written report cannot be a group effort and cannot be plagiarized from other sources.** The point values provided below are used as guidelines by the instructors and give you an indication of how each section is weighted. Use your book entitled *Writing Papers in the Biological Sciences*, 4<sup>th</sup> edition by Victoria McMillan as referenced below. Please **double space** your report. The title, your name, your section, and date do NOT need to be on a separate title page.

**Title** -1 point                   \* see pages 69 through 72 of McMillan  
\* make descriptive and centered on the work you performed in the lab  
\* keep concise (generally  $\leq 12$  words)

**Introduction** - 6 points       \* see pages 76 through 78 of McMillan  
\* provide background information on the biological context for your work, citing previous work in the field  
\* cite reference sources as shown on pages 137-144  
\* explain your objective(s) and/or hypotheses

For this assignment, the introduction should be approximately 1-1.5 pages. The purpose of the introduction is to tell the reader the larger context within which your research is set. For this assignment, the larger context is the need for biofuels and the use of genetic engineering to produce them. The introduction should draw the reader into the paper and spark interest in the topic, and the larger context should hone into the specific work performed in your study. You have had some practice in writing scientific objectives in a concise way. The objectives state the specific scope of the present study and clearly, but briefly, describe the methods used. The introduction gives the broader context of why the work you performed contribute to the scientific knowledge in the field. You are expected to find relevant literature regarding your topic; these sources must be cited and included in your 'literature cited' section. **Use the CSE style and name-year system for citations** (McMillan pages 138-144).

**Materials & Methods** – 6 points   \* see pages 78 through 83 of McMillan  
\* write complete sentences rather than lists of materials or steps  
\* be comprehensive; others must be able to repeat your procedure  
\* should be divided into sections with subheadings if lengthy  
\* use past tense and passive voice; tell us what you (or your lab partner) did, but leave yourself out of it.

The materials and methods section of the report should accurately describe the work that you performed in the lab, with any modifications duly noted. Include specific reagents and bacterial strains used, concentrations of solutions and all pertinent protocols. Remember your audience – your audience knows how to use the equipment in the lab, but may need to know the names of specialty instruments to replicate your work or evaluate it critically. For example, you don't need to describe the details of using SigmaPlot, but you should indicate that it was used and for what. Be concise!

**Results** - 6 points               \* see pages 83 through 89 and Chapter 3 of McMillan  
\* describe the data, emphasizing important trends and patterns;  
provide explanatory details with references to any tables or figures

- \* summarize your data using clearly labeled tables and/or graphs (note examples in Chapter 3)
- \*use past tense
- \* avoid the temptation to analyze or interpret results; that is saved for the discussion section

The results section will present the data that you obtained (a table, a figure and a graph). Clearly describe the results in the context of the controls performed in these experiments. Make sure the figure legends are descriptive and concise. The questions included in the lab above are meant to help you with this.

**Literature Cited** – 1 point      \* use the CSE style