Solutions to 7.014 Problem Set 7

Question 1

On your trip to the Amazon you were introduced to a new plant that the indigenous people use as an anti-viral treatment. You took samples home to your lab and you found that this plant makes a protein (the PV protein) that prevents viral replication. Excited by the possible anti-AIDS applications, you construct a genomic DNA library from this plant in the hope of cloning the *PV* gene.

a) What is a genomic DNA library?
A collection of bacterial cells (or phage particles) each of which contain a different small piece of the genomic DNA that you are interested in.

b) You get a DNA sample from cells, digest it with a restriction enzyme, and clone it into a vector. List 3 features of the vector that are absolutely required for your library construction.
An origin of replication (ori)
A cloning site
A selectable marker

c) Circle on the following lists ALL you would need in order to construct the genomic DNA library. Assume you start with intact plant genomic DNA.

Enzymes	Reagents
(Restriction enzyme)	Size separation gel
(Ligase)	Okasaki fragments
DNA Polymerase	ATP, TTP, CTP, GTP
RNA Polymerase	ddATP, ddTTP, ddCTP, ddGTP
Transcriptase	Primers
Reverse Transcriptase	Replication fork
3' to 5' exonuclease	(*E. coli* (bacteria))
(Cloning vector)	Human cells
	Virus

Briefly describe the function of each item circled.
Restriction enzyme: cut the vetor and the genomic DNA
Ligase: join together the cut DNA fragments
Cloning vector: receives the cut genomic DNA and allows propagation in the bacterial cell host.
E. coli: Acts as a host that receives the recombiant plasmids and replicated this new DNA

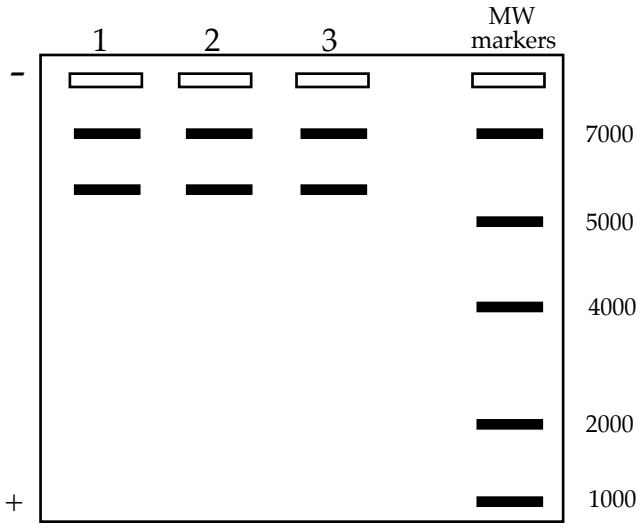
Question 1, continued

d) Your genomic DNA library is now ready. You have PV protein sequence information and want to make a DNA probe that is 12 nucleotides long. Given the protein sequence below, circle the amino acids that you will use to design your DNA probe. Why did you choose these amino acids?

leu-arg-met-gly-ser-ala-val-leu-pro-thr-arg-ser-trp-met-tyr-lys-arg-his-thr-ser-ala-cys-met-pro-phe

These amino acids can be encode by fewer possible DNA sequences that any other four consecutive amino acids.

You screen the library to find the *PV*gene. You find three different clones that hybridize to your probe. You purify the plasmid from each clone, digest it with the restriction enzyme you used when making your library, and then run the samples in a size separation gel. You observe the following pattern:

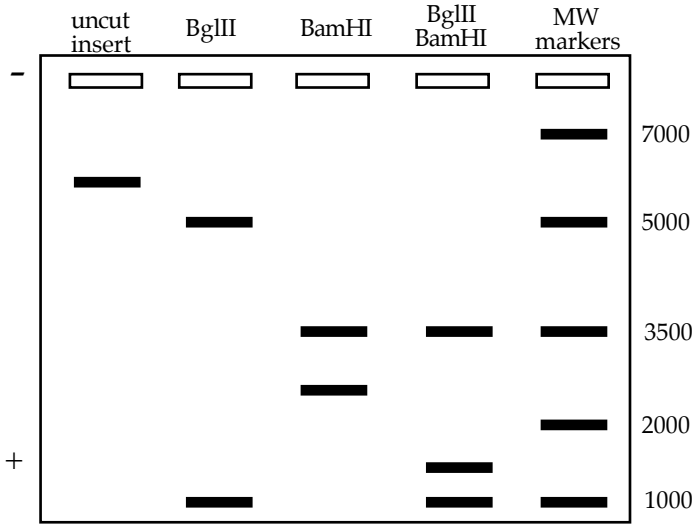


e) Why do you get two bands of different sizes for each clone?
The digest as described give one band for the vector, and one band for the insert.

f) Do you expect your gene to be inserted in the same direction in clones 1, 2 and 3? Why?
No, because we cut with only one enzyme the ends of the fragment are identical and it can insert in either direction.

Question 1, continued

You decide to analyze clone 1 further. You purify the genomic insert and then digest it with *Bgl*II and with *Bam*HI and then run the cut DNA on a size separation gel. You obtain the following results:

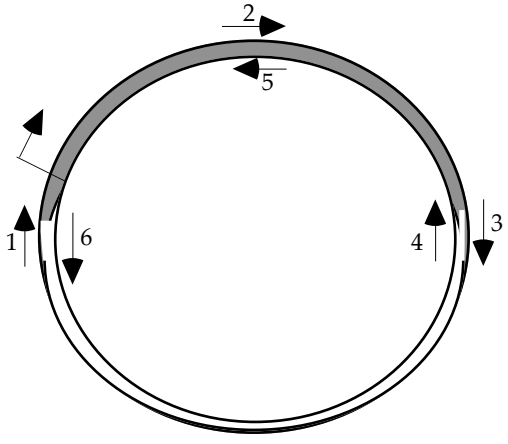


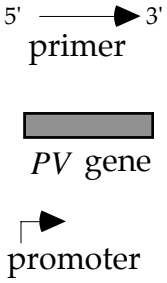
g) Draw a map of the genomic insert indicating the restriction sites for the enzymes *Bgl*II, *Bam*HI and the one you used to make the library.

h) You decide to sequence the 5' end of the *PV* gene in clone 1, and to PCR amplify the entire *PV* gene from clone 1. Sequencing reactions only provide about 600 bp of sequence. Choose the MINIMUM number of primers necessary for your experiments. Give the primer number(s) in the space provided.

For PCR: *1 and 4*

For sequencing: *1*





Question 2

You have discovered a new restriction enzyme, which you have decided to name NewI. Unfortunately all you know about this enzyme is that it is a restriction endonuclease. You decide to give this enzyme a variety of DNA sequences in order to determine the site at which it can cut.

a) You find that when long stretches of different sequences are cut by this enzyme it tends to cut once every 256 base pairs. Given that DNA is composed of four different bases (ACGT) and that NewI cuts at only one sequence, what is the most likely length of the sequence that NewI recognizes.

4^4 = 256
So it would cut every four base pairs.

b) You would now like to determine what exact sequence is recognized by NewI. You have a collection of DNA molecules that NewI cuts and a collection of DNA molecules that it fails to cut. What technique would you use to determine the recognition site?

Sequence all of the fragments that were cut (or alternatively all the fragments that were not cut) and look for a common sequence of four base pairs in all of them.

c) Once you have determined the sequence of the recognition site of NewI, you decide to use it to create a new vector. You have a vector that produces the bacterial protein β-galactosidase, and you want to modify this vector so that you can now produce human insulin. You successfully remove the DNA encoding the β -galactosidase protein (the promoter and the operator region remain) and replace it with the cDNA for the human insulin gene. You then transform the recombinant plasmid into *E. coli* cells. Your manipulations are successful, yet the cells carrying your new plasmid do not make human insulin. Give two (or more) reasons why your experiment failed.

The human cDNA may have inserted in the reverse orientation giving a protein that does not match the human insulin sequence. You also have to make sure that you induced the promoter using either lactose or a synthetic analogue such as IPTG. (Another problem is the fact that insulin needs to be post-translationally modified in order to be active and bacteria do not contain the enzymes to do this. In addition the lack of certain chaperones in bacteria will cause the human insulin to be mis-folded and found in inclusion bodies.)

Question 3

Combinatorial gene rearrangements occur within the cells that produce antibodies.

a) What advantage does this provide the immune system in terms of antibody production?

It allows the immune system to produce a huge diversity of antibodies without the need for individual genes for each antibody.

b) During the process of gene rearrangement heavy and light chains are formed from combination of V, D and J gene segments. Assume that:

(1) any V_H segment can join to any J_H or D_H segment and any V_L segment can join to any J_L segment
(2) any D_H segment can join to any J_H segment, and
(3) any light chain can associate with any heavy chain.

Given the following genome organizations, determine how many different antibody molecules can be constructed.

(i) L: 8 V_L and 4 J_L
H: 8 V_H, 4 D_H and 4 J_H
*8*4*8*4*4 = 4,096*

(ii) L: 100 V_L and 2 J_L
H: 100 V_H, 10 D_H, and 2 J_L
*100*2*100*10*2 = 400,000*

(iii) L: 2000 V_L and 4 J_L
H: 2000 V_H, 20 D_H and 4 J_H
*2000*4*2000*20*4 = 1,280,000,000*

Question 4

A common method for diagnosing an individual as being HIV(+) or HIV(-) is to look for antibodies to the virus within the serum of that individual. A positive result from this test generally indicates that an individual has been exposed to the virus. A negative result from this test does not necessarily mean that the individual is in fact not infected.

a) How could an individual be infected with HIV and yet test negative by the antibody test?

If you were only recently infected with the virus, your body would not yet have produced an immune response and so the test would be negative.

b) Given that HIV integrates itself into the genome and that its sequence is known, what other method could be used to detect the presence of the HIV virus? Briefly describe this technique, listing the reagents that are necessary to perform it. (*Hint: The amount of viral DNA in one human cell is not very great, how would you make more viral DNA).*

You could do PCR using viral specific primers. You would need T-cell DNA, Taq polymerase, buffer, dNTPs, and the primers. You would denature the T-cell DNA by heating, anneal the primers by cooling and then amplify the DNA sequence using the polymerase. You repeat this technique 30-40 times to produce massive quantities of viral DNA. If you run a gel of the PCR and see a band it would tell you that there was viral DNA present in the T-cells that you tested.

	U		C		A		G		
U	UUU	**phe (F)**	UCU	**ser (S)**	UAU	**tyr (Y)**	UGU	**cys (C)**	U
	UUC	**phe**	UCC	**ser**	UAC	**tyr**	UGC	**cys**	C
	UUA	**leu (L)**	UCA	**ser**	**UAA**	**STOP**	**UGA**	**STOP**	A
	UUG	**leu**	UCG	**ser**	**UAG**	**STOP**	UGG	**trp (W)**	G
C	CUU	**leu**	CCU	**pro (P)**	CAU	**his (H)**	CGU	**arg (R)**	U
	CUC	**leu**	CCC	**pro**	CAC	**his**	CGC	**arg**	C
	CUA	**leu**	CCA	**pro**	CAA	**gln (Q)**	CGA	**arg**	A
	CUG	**leu**	CCG	**pro**	CAG	**gln**	CGG	**arg**	G
A	AUU	**ile (I)**	ACU	**thr (T)**	AAU	**asn (N)**	AGU	**ser (S)**	U
	AUC	**ile**	ACC	**thr**	AAC	**asn**	AGC	**ser**	C
	AUA	**ile**	ACA	**thr**	AAA	**lys (K)**	AGA	**arg (R)**	A
	AUG	**met (M)**	ACG	**thr**	AAG	**lys**	AGG	**arg**	G
G	GUU	**val (V)**	GCU	**ala (A)**	GAU	**asp (D)**	GGU	**gly (G)**	U
	GUC	**val**	GCC	**ala**	GAC	**asp**	GGC	**gly**	C
	GUA	**val**	GCA	**ala**	GAA	**glu (E)**	GGA	**gly**	A
	GUG	**val**	GCG	**ala**	GAG	**glu**	GGG	**gly**	G