NAME_____TA_____ SEC____

7.013 Problem Set 3 FRIDAY March 5, 2004

Problem sets will NOT be accepted late.

Question 1

Shown below is the sequence of a short eukaryotic gene. Only the 5′ to 3′ sequence of the "RNA-like" or complementary strand of the template strand is shown.

5′– GATGAGTTAT[10] AATATTTCTC[20] TCCAGGCATG[30] GAGTATTCCG[40]

GTGTGCGATC[50] TCCCCCTTTG[60] GACCACCCTG[70] GGTTGCCCTC[80]

TAAGCATAAT[90] AGTTGGCCAT[100] ACGTTTCTGT[110] AATTAAAATT[120]

TGTTTGCCTC[130] ATGT –3′

This is what's known about this sequence…

- RNA polymerase and associated proteins recognize the sequence 'TATAAT' and initiate transcription five nucleotides downstream of the sequence at position 18.
- The intron splice sites have been determined to be 'CUU' (5′ splice site) and 'AAG' (3′ splice site). These sequences and the sequences between these sites will be spliced out as introns in the processing towards mature mRNAs.
- The organism cleaves the RNA and adds poly-A tails immediately following the sequence 'AGUUGG.' These tails are 14 nucleotides long. RNA is produced and processed in the following order: transcription, splicing, polyadenylation, and 5′ capping.

a) What is the sequence of the first 20 RNA base pairs of mRNA? Denote 5′ and 3′ ends.

CUCUCCAGGCAUGGAGUAUU

b) What is the sequence of the processed, cytoplasmic-localized mRNA produced from this gene?

5′cap– CUCUCCAGGCAUGGAGUAUUCCGGUGUGCGAUCUCCCCCAUAAUAGUUGGAAAAAAAAAAAAAA 3′

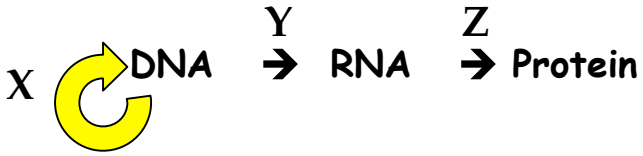
c) What is the sequence of the short protein encoded by this gene? Put the answer in the table below next to "original sequence". (Use the single letter code from chart at end.) What would the protein sequence be if the following point mutations occurred?

	Protein	Type of mutation choose from frameshift, nonsense, missense, silent
Original sequence	MEYSGVRSPP	
mutation of base 33 (**G mutated to T**)	MDYSGVRSPP	missense
mutation of base 36 (**T mutated to A**)	ME	nonsense
mutation of base 51 (**T mutated to C**)	MEYSGVRSPP	silent
mutation of base 70 (**G mutated to C**)	MEYSGVRSPP	In intron (~silent)

Question 2

Below is a diagram of the "central dogma." For each step (X, Y, Z) fill in the following:

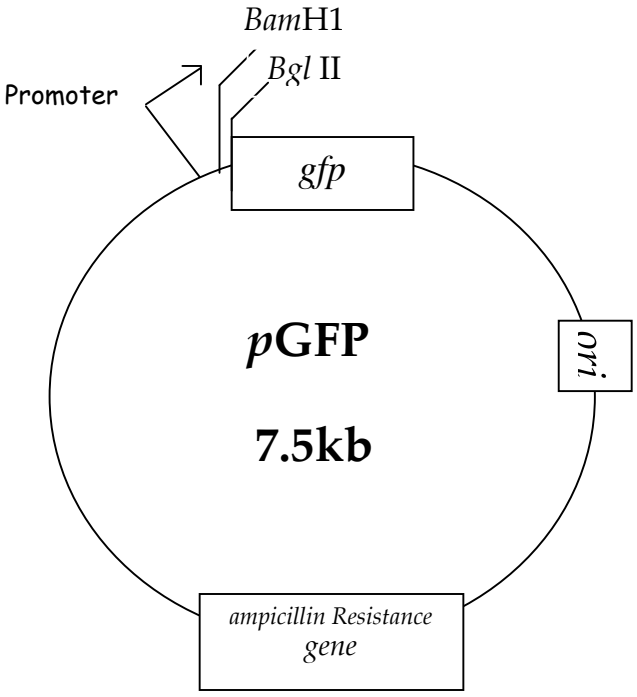
a) Name of the process
b) Name of the primary enzyme or enzyme complex
c) Template Macromolecule – the template read by the enzyme or enzyme complex
d) Monomers polymerized –to form the new macromolecule
e) Initiation site – the name of the sequence in the template that directs the enzyme to begin polymerization.



	X	Y	Z
a	DNA Replication	Transcription	Translation
b	DNA Polymerase	RNA Polymerase	ribosome
c	DNA	DNA	RNA
d	Deoxyribonucleoside triphosphates	Ribonucleoside triphosphates	Amino Acids
e	Origin of replication	promoter	AUG, start codon

Question 3

You've gained a position as a technician in a prestigious yeast lab. Your first assignment is to learn the cellular localization of product of the gene, *bio701*, studied by the lab. One exciting technique involves creating a protein fusion where your protein of interest is fused with GFP, green fluorescent protein. The resulting fused protein will fluoresce green when excited at the appropriate wavelength. To make this fusion you must fuse the *bio701* gene with the *gfp* gene. Certain vectors have been made that simplify the construction by allowing you to insert your gene in frame with the gfp gene. See below.



You know that the *Bgl* II site sequence in pGFP is in frame with the *gfp* gene (such that the AGA will encode arg (R) and the TCT will correspond to ser (S)). (see below)

The *bio701* coding region is about 1500 bases long.
The actual sequence corresponding to proximal and distal ends of corresponding to the coding region *bio701* is shown below. The middle 1.4 kb are not shown.

```
5'ATGaccatgggcgacaagaagagcccgaccaggcc...
                                (1.4kb of internal base pairs)
                                                ...gggcgacatgtcagcagtcaatgatgaatcttTGA 3'
```

The first ATG corresponds to the start codon for *bio701*. The capitalized TGA at the end corresponds to the stop codon of *bio701*.

Since your goal is to make a protein fusion where the N terminus of Green Fluorescent protein is replaced by the Bio701 protein, you must engineer a DNA construct that fuses the *bio701* DNA in frame with the *gfp* gene sequence.

And since you must design primers 20 nucleotides long to amplify *bio701* from yeast genomic DNA you cleverly decide to engineer restriction sites at the end of these primers that will allow you to cut the resulting amplified product with *Bam* H1 and *Bgl* II, and insert it into the *Bam* H1 *Bgl* II sites of *p*GFP. (You've checked that there are no *Bam* HI or *Bgl* II sites within *bio701*.) These restriction sites are given below. The arrows represent where the enzyme will cleave the DNA.

So first you design a primer to amplify *bio701* in one direction with an engineered *Bam* H1 site attached at the 5' end such that you will be able to cut the final amplified PCR product with *Bam* H1 enzyme to insert into the *Bam* H1 site of *p*GFP. Your primer looks like…

Forward primer: 5' ggatccatgaccatgggcga 3'

Your boss quickly looks at your design and approves.
Design a primer to amplify the reverse direction with an engineered *Bgl* II restriction site that will allow you to digest the PCR product and then ligate it into the *Bgl* II site **in frame** with the *gfp* gene.

Reverse primer: 5' agatctaagattcatcattg 3'

Assuming your primers work, fill in the blanks to outline a working cloning strategy.

PCR amplify the *bio701* gene.

Cut PCR product (or *bio701*) with *Bam* H1 and *Bgl* II

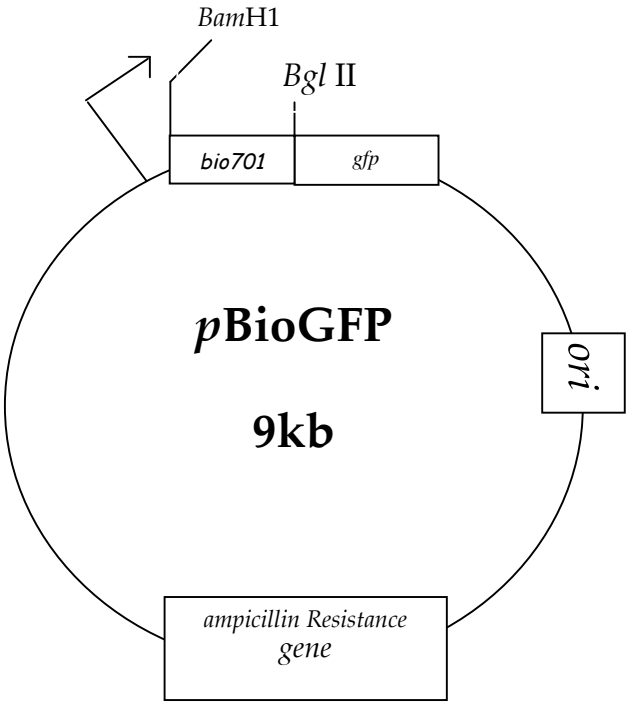
Cut *pGFP* (or vector) with *Bam* H1 and *Bgl* II

Ligate gel-purified cut fragments

Transform bacteria

Select for Transformants on Ampicillin plates

Your desired construct should look something like this:



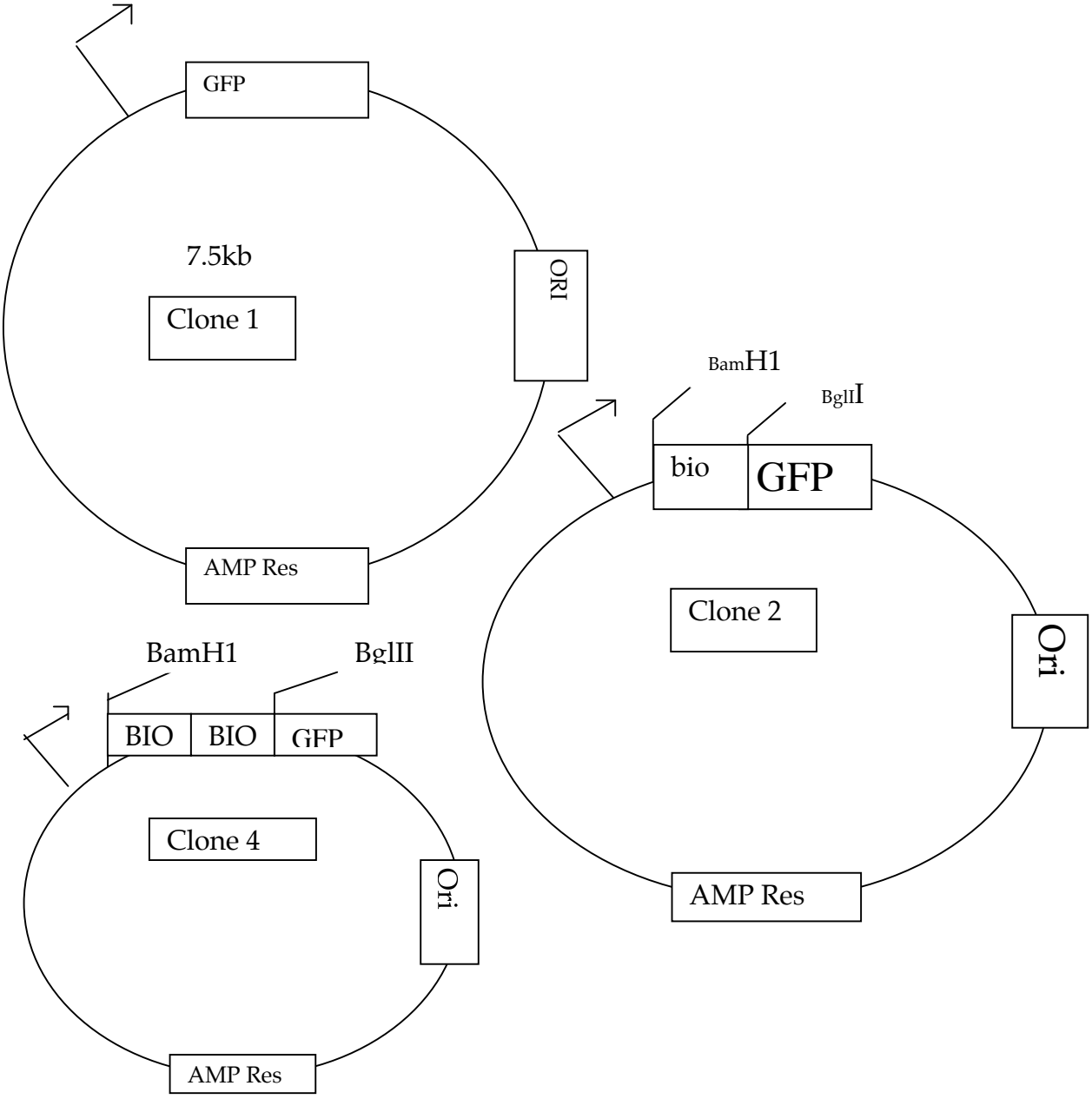
After you have obtained bacterial transformants, you wish to isolate their plasmids to screen and determine that everything went correctly with the ligation. You pick 4 bacterial colonies and harvest the plasmid clones from them and perform restriction digests. Then you perform gel electrophoresis with the restriction digests. The results are as follows:

	Bam H1	*Bgl* II	*Bam* H1 + *Bgl* II
Clone 1	(uncut) ~7.5kb	(uncut) ~7.5kb	(uncut) ~7.5kb
Clone 2	(uncut) ~9.0kb	(uncut) ~9.0kb	(uncut) ~ 9.0kb
Clone 3	9.0kb	9.0kb	7.5kb, 1.5kb
Clone 4	10.5kb	10.5kb	7.5kb, 3.0kb

Which, if any, of the clones could possibly be the right construct?__ Clone 3_____

Some of the clones are obviously not the desired construct. Draw plasmid maps for what these constructs might look like, and give a plausible explanation for what may have gone wrong.

Clone 1: *Bam* H1 and *Bgl* II ends are compatible, no insert –they ligate together and cannot be recut
Clone 2: Since *Bgl* II and *Bam* H1 are compatible ends, the insert went in the wrong direction, but destroyed *Bam* H1 and *Bgl* II sites so that they are no longer the same, and the resulting plasmid can't be recleaved by these enzymes.
Clone 4: Two inserts were ligated into into the vector

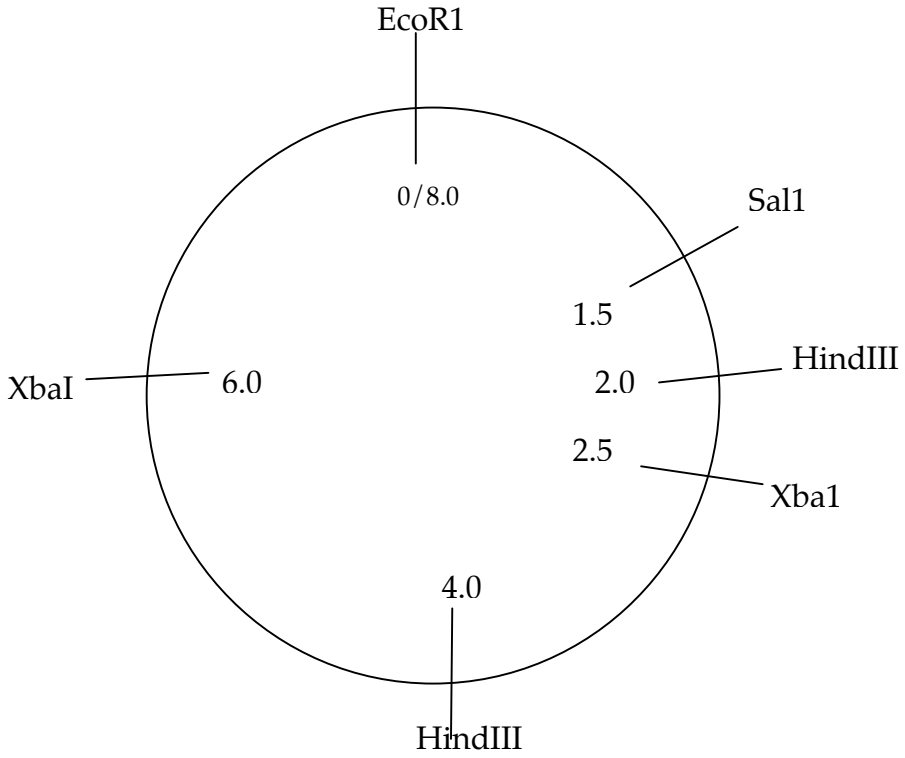


Question 4

You are given a plasmid that contains 8kb of DNA. You wish to create a restriction map of this plasmid by performing a series of restriction digests. After the digests are complete, you perform electrophoresis on an agarose gel and after staining the gel to visualize the DNA you measure the fragment sizes.

Restriction enzymes in reaction	Size of DNA fragment(s) in kilobases
Eco R1	8.0
Hind III	6.0, 2.0
Sal I	8.0
Xba 1	3.5, 4.5
Eco R1, *Hind* III	4.0, 2.0, 2.0
Eco R1, *Sal* 1	1.5, 6.5
Eco 1, *Xba* 1	2.0, 3.5, 2.5
Hind III, *Sal* I	0.5, 2.0, 5.5
Hind III, *Xba* 1	0.5, 1.5, 2.0, 4.0

Draw a restriction map of this plasmid with the *Eco* R1 site listed as position 0/8.0kb. Indicate the position and distance of the other restriction sites based on the above analysis.



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