Assignment 1

Note: Part I of the assignment is to be submitted through Blackboard (See instructions on Blackboard). Parts II, III and IV are to be submitted as a hard copy during your lab session.

Part I

Dilutions and Concentrations

When doing your calculations, do not round off your intermediate numbers. Only round off the final answer. Your answers must be submitted to two significant figures after the decimal. For example 2.00, 0.020, or 0.0020. It is strongly recommended that you submit your answers using the web browser Firefox. **You will be allowed two submissions!**

Use the following information to answer questions 1-6

You prepare a solution by adding the following ingredients in the order indicated: 600 mL H_2O, 125 mL 1.6 M LiCl, 50 mL 20 % (m/v) $MgCl_2$, and 25 mL 10g/L NaCl. The properties of each ingredient are as follows:

LiCl: MW 200g/mole, density 1.2g/mL, density of 1.6 M soln. 1.05g/mL

$MgCl_2$: MW 150g/mole, density 1.3g/mL, density of 20% soln. 1.1g/mL

NaCl: MW 35g/mole, density 1.15g/mL, density of 10g/L soln. 1.03g/mL

Final solution: density: 1.25g/mL

1. What is the final molarity of $MgCl_2$ in the solution? (**0.083M**)
2. What is the volume in milliliters of one part? (**25 mL**)
3. What is the percentage (m/m) of NaCl in the final solution? (**0.025%**)
4. What is the percentage (m/v) of LiCl in the final solution? (**5%**)
5. What is the number of parts of solvent in the final solution? **(24 parts)**
6. What is the percent of total solute (m/v) in the final solution? (**6.28%)**

7. You start with 2.5L of a KNO_3 stock solution and wish to prepare 10.0 L of 1.5 M KNO_3. What percentage (m/v) would the potassium nitrate stock solution need to be if you were to use it all? (MW of KNO_3: 101g/mole) **(60.6%)**
8. You have a 5.0 M copper sulfate stock solution and a 2 M NaCl stock solution. You wish to prepare a solution with a final concentration of 0.25 M copper sulfate and 0.25 M NaCl containing 330 mL of water as a solvent. How many milliliters of the NaCl stock solution would the solution contain? (**50 mL**)
9. 40.0 mL of 2.0 M $Fe(NO_3)_3$ is mixed with 2 mL of 5 M $Fe(NO_3)_3$ and 48 mL of water. What is the final molar concentration of NO_3^- ions? (**3M**)
10. You add 3.5 L of an HCl solution of unknown concentration to 2.0 L of 0.5 M HCl and 4.5 L of water. The final concentration of HCl was 1.5 M. What was the unknown concentration of the initial HCl solution? (**4M**)
11. You prepare 100 mL of a 20% NaCl (m/v) solution. (Denisty of NaCl: 2.16 g/mL). How many milliliters of water are in this solution? (**90.74mL**)

12. What is the molar concentration of chloride ions in a solution prepared by mixing 100.0 mL of 2.0 M KCl with 50.0 mL of a 1.50 M $CaCl_2$ solution? (**2.33M**)
13. The A_{260nm} of a DNA solution is 0.15. How much of this DNA solution and a 5.5X loading dye should you add to 20 µL of water to obtain a sample which contains 75 ng of DNA in 0.5X loading dye? (A_{260nm} of 1.0 = 50 µg/mL DNA) (**10 µL DNA and 3 µL Dye**)
14. Solution A has 1.20 mg of protein per mL. Solution B has 3.10 mg of protein per mL. If you combine 34 mL of solution A with 19 mL of solution B, what is the protein concentration in µg/mL of the final solution? (**1881.13 µg/mL**)
15. You have a 0.35M solution of glycine (MW: 70g/mole). By what factor must this solution be diluted to obtain a final concentration of 0.49% (m/v)? (**5**)

Restriction Enzymes & Restriction Mapping

16. The nomenclature of restriction enzymes can provide useful information about the source of the enzyme. For example, *Eco*RI indicates that this enzyme was the first enzyme isolated from an *Escherichia coli* strain "R". From which bacterial genus was *Bgl*II isolated from?
    Bacillus
17. Define the following terms: Isoschizomer, neoschizomer, and isocaudomer.
    Isoschizomer: Pairs of restriction enzymes specific to the same recognition sequence and which generate the same termini.
    Neoschizomer: Pairs restriction enzymes that recognize the same nucleotide sequence but cleave at a different site generating different termini.
    Isocaudomer: Pairs of restriction enzymes with different recognition sequences but generate identical termini.
18. Amongst the enzymes listed below, which if any, generate compatible ends to each other? (Ex. A and B)

	Enzyme	**Recognition Sequence**
A	*AccI*	GT/CGAC
B	*ClaI*	AT/CGAT
C	EagI	C/GGCCG
D	*Taq*I	T/CGA
E	*NsiI*	ATGCA/T
F	*NotI*	GC/GGCCGC
G	*PstI*	5CTGC/AG

A+D
A+B
B+D
C+F

19. The partial sequences recognized by two restriction enzymes "A" and "B" are indicated below. Complete the sequences such that palindromes are generated for each of the sites. On your palindromes indicate which phosphodiester linkage would have to be cleaved such that protruding ends of enzyme "A" are compatible with protruding ends of enzyme "B". **Ex. CT/GCAG**

    Enzyme "A": 5'CTAATTAG 3'

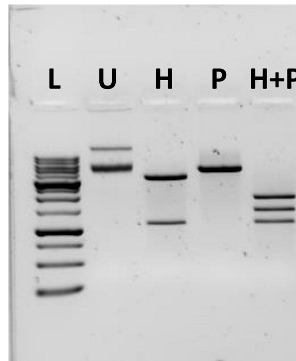
    Enzyme "B": 5'AAA TTT 3'

20. Consider your answer to the previous question. A DNA fragment generated with enzyme "A" was ligated to a DNA fragment generated with enzyme "B" as illustrated below. Which enzymes, A, B, A & B, or neither A or B would cut the ligated DNA?

———————A/B———————

CT/AATTAG A/AATTT
CTA/ATTAG AA/ATTT
CTAATT/AG AAATT/T
CTAAT/TAG AAAT/TT

21. This picture represents an agarose gel electrophoresis of various restriction digests of the plasmid pBR322. (This file can be obtained from this course's web site under the heading "Data>pBR322).



L: Molecular weight ladder
U : Undigested pBR322
H : pBR322 digested with *Hinc*II
P: pBR322 digested with *Pvu*II
H+P: pBR322 digested with *Hinc*II and *Pvu*II

Answer the following questions based on the results obtained:

a. How many times did *PvuII* cut within the plasmid? 1
b. How many times did *Hinc*II cut within the plasmid? 2
c. How many times did *Hinc*II cut within the *PvuII* fragment? 2
d. How many different fragments of different sizes could be generated if the *Hinc*II + *Pvu*II digest was partial? 7

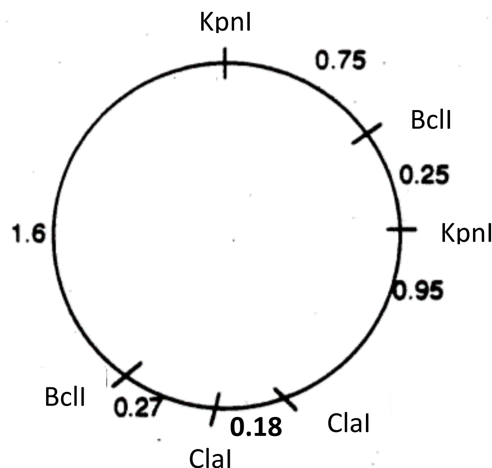
22. A linear 12 Kbp DNA **fragment** shown below has cleavage sites for *Bam*HI and *Eco*RI. The numbers indicate the distance in kilobases. Indicate what fragment sizes would be observed on an agarose gel following digests with *Bam*HI, *Eco*RI, and *Bam*HI + *Eco*RI. Note, if different fragments of the same size are generated, the size should only be indicated once. (For example do not indicate 2Kbp and 2Kbp)

23. What fragment sizes could be generated from a *Bam*HI partial digest? Only indicate the sizes of **intermediate** fragments which would not be obtained following a complete digest. (Fragments which contain one or more *BamHI* site which remains undigested) 6, 11
24. It was determined that the enzyme *Xho*I cuts at 2.0Kbp on the map shown above. Which *Bam*HI fragment size would be cut by *Xho*I? 5kb
25. A complete digest with *Eco*RI + *Bam*HI of 12µg of the above fragment was performed. Indicate the amount which would be obtained in µg of the 4 Kbp fragment. 4ug

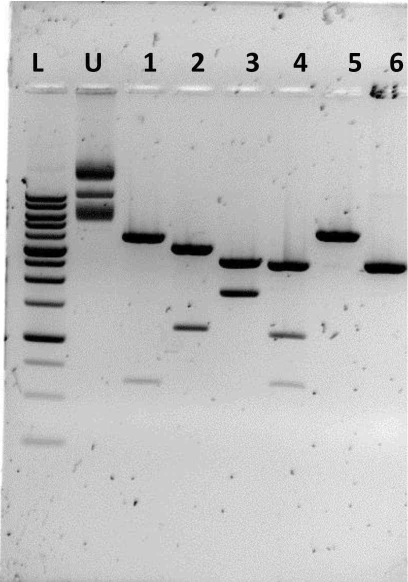
Part II: Restriction mapping continued (10 points/question)

1. Plasmid pDA401 (total size = 4.0 kb) was digested with the enzymes *Cla*I, *Bcl*I, and *Kpn*I. The results obtained in each digestion are presented below. From this data, construct a restriction map of pDA401.

Digest	**Size of Fragments**
ClaI	3.82 kb, 0.18 kb
*Bcl*I	2.35 kb, 1.65 kb
*Kpn*I	3.00 kb, 1.00 kb
ClaI + *Bcl*I	2.35 kb, 1.20 kb, 0.27 kb, 0.18 kb
ClaI + *Kpn*I	1.87 kb, 1.00 kb, 0.95 kb, 0.18 kb
*Bcl*I + *Kpn*I	1.60 kb, 1.40 kb, 0.75 kb, 0.25 kb



Lab exercise: Restriction digests & agarose gel electrophoresis (Pg. 18-20)
Below is shown an agarose gel electrophoresis of the predigested DNA as you performed in the lab. (This file can be obtained from the course's web site under the heading Data>predigests)



L: 1Kbp molecular weight markers
U: Undigested recombinant **pUC9** plasmid DNA
1: Recombinant **pUC9** plasmid cut with *Bam*HI
2: Recombinant **pUC9** plasmid cut with *Eco*RI
3: Recombinant **pUC9** plasmid cut with *Hind*III
4: Recombinant **pUC9** plasmid cut with *Eco*RI + *Hind*III
5: Recombinant **pUC9** plasmid cut with *Pst*I
6: **pUC9** vector cut with *Bam*HI

2. Submit a standard curve of the molecular weight ladder (Migration distance Vs. Size in Kbp)
   Grading: 2 points/criteria listed below
   Graph was computer generated
   Title provided
   Appropriate axis labels with units is provided
   Y axis is a logarithmic scale
   Trend line provided (may be drawn by hand) which has a good linear range

3. Submit a table presenting the results of the restriction digests of the **recombinant** plasmid which includes the following information: Enzyme used, Total number of cuts, Number of cuts in the insert, Number of cuts in the vector, fragment sizes observed. Note sizes are approximate.

Enzyme	Total cuts	Cuts in insert	Cuts in vector	Sizes in Kb
BamHI	2	0	1	4.5, 0.7
EcoRI	2	1	1	4.0, 1.2
HindIII	2	1	1	3.7, 1.8
PstI	2	1	1	5.3
EcoRI + HindIII	3	2	2	3.7, 1.0, 0.8

4. In a caption accompanying the table, indicate the total size of the plasmid, the size of the vector, the size of the insert, and the restriction site (s) in which the insert was introduced in the vector. Give points if the following information is provided:
   Total size in the range of 5.2 Kb
   Size of vector in the range of 3.7 Kb
   Size of insert in the range of 1.8Kb
   Insertion site: HindIII

5. Submit a figure which represents a possible restriction map of the insert within the multiple cloning site of pUC9. Your map should be linear, to scale and only include the insert within the multiple cloning site. (See directives on this course's web site under the heading graphs and figures)

Map can be computer drawn or hand drawn.
Should be to scale
Must include scale bar
Should indicate following regions: multiple cloning site and insert
Insertion site should be easily determined from figure

Part III: Bioinformatics 1 (10 points/question)

1. What is the protein accession number of the first record for actin, cytoplasmic 2 isoform 1 from the mouse (*Mus musculus*) obtained through a general search? NP_001300852.1
2. Does the record with the accession number NG_009024 correspond to a nucleotide or a protein record? nucleotide
3. What was the source of the sequence corresponding to the record with the accession number NG_009024; protein, mRNA, or genomic DNA? DNA
4. Submit a printout of a FASTA sequence of one of the unknown genes. Include the name of the gene as a heading to the printout. Give points if provided with gene name.
5. Submit the following information with regards to each of the unknown genes from the first bioinformatics exercise.

- **Accession number**
- **Coverage**
- **Ident.**
- **E value**
- **The definition**
- **The organism from which this sequence was obtained**
- **The gene name**
- **The gene's product name**
- **The protein's accession number**

Acc. Num.	Cov.	Ident.	E value	Definition	Organism	Product	Pro. Acc.
NM_057444.3	100%	100%	0.0	Drosophila melanogaster yellow (y), mRNA	Drosophila melanogaster	yellow	NP_476792.1
X91249.1	100%	100%	0.0	H.sapiens mRNA for white gene protein	Homo sapiens	white	CAA62631.1
NM_001110389.1	45%	91%	0.0	Danio rerio mitogen-activated protein kinase 8a (mapk8a), mRNA.	Danio rerio	mitogen-activated protein kinase 8	NP_001103859.1
Y00417.1	42%	100%	0.0	Wheat mitochondrial COI gene for cytochrome oxidase subunit I	Triticum aestivum	cytochrome oxidase subunit I	CAA68474.1
XM_009805932.1	74%	90%	0.0	Nicotiana sylvestris uncharacterized LOC104249497	Nicotiana sylvestris (wood tobacco)	LOC104249497 isoform X1	XP_009804234.1

Part IV: Lab performance (10 points each)

Dilutions exercise with micropipettors (Pg. 13)

1. Indicate the absorbance readings obtained for each of the following solutions which you prepared during lab exercise 1.

    a. A 0.2mM solution of compound "A".
    b. A 0.036% (m/v) solution of compound "B".
    c. A 5% (v/v) solution of solution I.
    d. A solution containing 0.1mg of compound "A" and 0.1% (v/v) of compound "B".
    e. A solution with the following ratio: solution I: solution II : water : 2 : 1 : 247

Grading: Within 20% of values provided: 100%
Between 21-50% of values provided: 50%
Greater than 50% difference: 0%

Determining the concentration of DNA: (Pg. 15)

2. Submit a graph representing the A260 readings Vs standard DNA concentrations. Include a line of best fit, the R coefficient, and the formula of the line on the graph.
   Includes required information
   Axes are appropriately labeled
   Title is included
   R value is 0.9 or higher

3. According to your graph what was the DNA concentration of the undiluted unknown DNA solution provided?

Within 20% full points
Between 21%-50% 50% of points
Deviation from expected value greater than 50% 0 points.

Restriction digests & agarose gel electrophoresis (Pg. 16-20)

4. Submit an appropriate figure of your gel electrophoresis including an accompanying legend. (Follow the directives for figures and graphs available on this course's web site)

   (One point/criteria)
   Figure provided with Accompanying legend
   Plasmid isolation (lane 2) got something of acceptable yield
   Figure title provided
   Lanes are labelled and easy to understand
   First lane is mw ladder
   Second lane is undigested control
   MW sizes of standard are indicated
   Loading is consistent between lanes
   Gel clearly shows bands and migration is satisfactory
   Legend indicates parameters of migration: Voltage, agarose concentration, samples loaded

Assignment #2

Note: Part I of this assignment is to be submitted through Blackboard (See instructions on Blackboard). Parts II, III and IV are to be submitted as a hard copy during your lab session.

Part I: Restriction digests and mapping

The table below presents the results of different digests of a plasmid. All sizes are in base pairs.

*Bam*HI	13199, 8572, 4627
*Kpn*I	13199
*Nhe*I	13199
*Bam*HI +*Kpn*I	12126, 9645, 8572, 3554, 1073
*Bam*HI + *Nhe*I	10701, 11070, 8572, 2498, 2129
*Nhe*I + *Kpn*I	11774, 1425

1. One of the enzymes used only partially digested the DNA. Which enzyme is it? BamHI
2. Indicate the size in base pairs of one of the intermediate products which represents an incompletely digested DNA fragment in the double digests. 12126, 9645, 10701
3. What is the total plasmid size in base pairs? 13199
4. What are the distances between the *Kpn*I and *Nhe*I restriction sites? 1425 + 11774
5. What are the distances between the *Bam*HI and *Nhe*I restriction sites? 2498 + 2129
6. What are the distances between the *Bam*HI and *Kpn*I restriction sites? 1073 + 3554

The table below presents the results of different complete digests of a linear DNA fragment. All sizes are in base pairs.

*Nco*I	5023, 1295
*Nru*I	4229, 2089
*Xba*I	3242, 2374, 702
*Nco*I + *Nru*I	2934, 2089, 1295
*Nco*I + *Xba*I	3242, 1295, 1079, 702
*Nru*I + *Xba*I	2374, 1855, 1387, 702

7. What is the total size of the undigested fragment? 6318
8. Among the fragments obtained with the *Xba*I digest, which one has *Xba*I overhangs on both ends? Indicate the size of the fragment. 3242
9. Digestion of the fragment indicated in question 8 with *Nru*I would generate fragments of what sizes? 1855, 1387
10. Among the fragments generated in the *Nco*I + *Nru*I digest, which one could be ligated and cloned in a vector digested with *Nco*I + *Nru*I? Indicate the size of the fragment. 2934
11. What is the distance between the *NcoI* and *NruI* restriction sites? 2934
12. What are the distances between the *XbaI* sites and the ends of the fragment? 2374 & 702

13. The restriction enzyme *Apo*I cleaves the sequence R/AATTY (R= A or G and Y = C or T). How many different palindromes does *Apo*I recognize? 2

14. What would be the average fragment size which would be expected following the digestion of genomic DNA with *Apo*I? Assume an equal distribution of A, G, C, and T. 2048

15. True or false; an overhang generated following the cleavage of a palindrome recognized by *Apo*I would necessarily be compatible with any other palindrome cleaved by *Apo*I? true

16. The enzyme *Psp*N4I cleaves the sequence GG/NNCC (N= any of the four bases). How many times would you expect *Psp*N4I to digest a 30 kb genome? Round off to the nearest whole number.29

17. True or false, an overhang generated following the cleavage of a palindrome recognized by *Psp*N4I would necessarily be compatible with any other palindrome cleaved by *Psp*N4I? False

18. Sequence A, which contains two *Bst*BI sites (TT/CGAA), was digested with *Bst*BI. The resulting fragment was then ligated into the unique *Taq*I restriction site (T/CGA) within vector B.

    Sequence A: CAG **TT/CGAA** TTC • • • • • • • • • • • • • • • • GGC **TT/CGAA** AAG

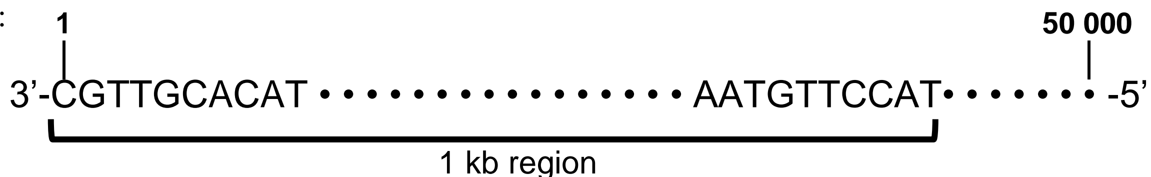
    Vector B: TGG **T/CGA** CAC

    Which enzyme (s) could be used to release the cloned DNA fragment from the recombinant vector B? *Bst*BI, *Taq*I , both, or neither. TaqI

PCR

Consider the following information to answer questions 19-24:

The partial sequence of a 1 kb region at the end of a linear single stranded 50 kb genome is shown:



19. Indicate the sequence of a **forward** primer of ten bases that could be used in a PCR amplification to amplify the 1 kb region. TACCTTGTAA

20. Indicate the sequence of a **reverse** primer of ten bases that could be used in a PCR amplification to amplify the 1 kb region. GCAACGTGTA

21. What is the minimum number of cycles required to obtain a double stranded amplification product which starts and ends with the primer sequences you've designed? 3

22. PCR amplification was performed with 10ng of the above described genome using the primers you've designed. What mass of the double stranded PCR product (in µg) would you have after 30 cycles? 53687ug

23. How many **additional** PCR cycles would be required to attain the same yield of product which was obtained in the previous question if you had started with 1 ng of genomic DNA instead? 3.3

24. Shown below, is the partial sequence of the region of the genome described previously:

1 ... 50 000

3'- **CGTTGCACAT** TTAAGGCCACGTGC• • • • • • • • • • • • • CCCGGGATATTT **AATGTTCCAT** AGAGCTCTGTGA• • • • • -5'

1 kb region

Indicate the sequence of the first 20 bases of the PCR product which would be synthesized using your forward primer.

TACCTTGTAATTTATAGGGC

GCAACGTGTAAATTCCGGTG

25. Your primers allowed you to amplify the 1 kb region of interest. You now wish to clone the amplified product in a 2.5 kb vector. To do so, you digest the amplicon and 100ng of the vector with the appropriate enzymes. How much of the digested amplicon (in ng) should be added to the ligation mix in order to have an insert to vector ratio of 3:1? 120ng

Part II: Lab exercise: Restriction mapping of a recombinant plasmid (Pg. 25-31)

Below is shown an agarose gel electrophoresis of the digests of the unknown recombinant pUC19 plasmids (A: orientation 1, B: orientation 2) as you performed in the lab. (This file can be obtained from the course's web site under the heading Data>pUC19 recombinant)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

1. mol. wt. marker
2. A uncut
3. A Bam HI
4. A Hin DIII
5. A Pst I
6. A Xba I
7. A Hin DIII + Pst I
8. A Hin DIII + Xba I
9. A Pst I + Xba I
10. pUC19 Bam HI
11. mol. wt. marker
12. B uncut
13. B Bam HI
14. B Hin DIII
15. B Pst I
16. B Xba I
17. B Hin DIII + Pst I
18. B Hin DIII + Xba I
19. B Pst I + Xba I
20. pUC19 Bam HI

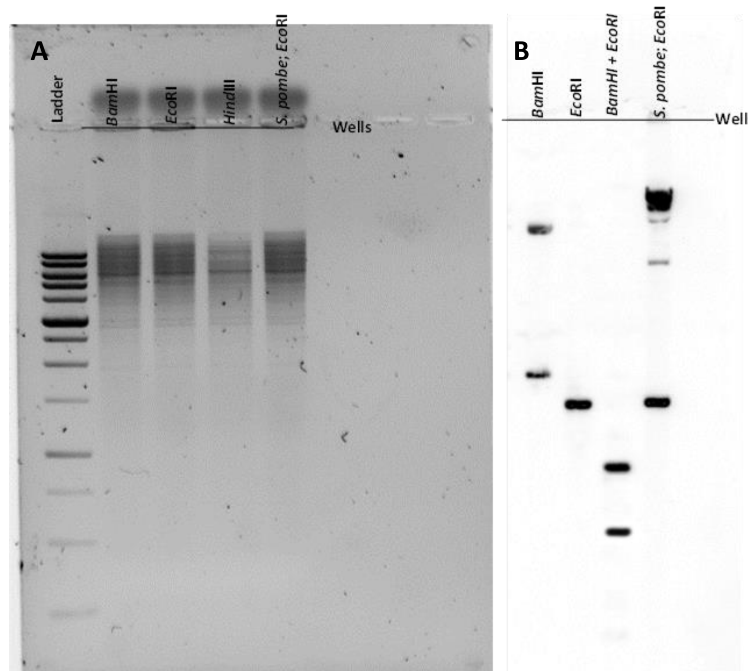
1. Submit a table presenting the analysis of the restriction digests. Your table should include: Enzyme (s) used, Total number of cuts, Number of cuts in the vector, Number of cuts in the insert, and Fragments sizes generated.

2. Which lanes, if any, show evidence of a partial digest? Indicate the sizes of any intermediate fragment and how many sites remain undigested.

3. In which restriction site of the multiple cloning site was the insert introduced?
4. What is the size of the insert?
5. Submit a figure of the restriction maps of the insert. Your maps must be linear, include the multiple cloning site, the insertion site, the size of the insert, the positions in the multiple cloning site or the insert of all the enzymes tested. Your figure must be to scale. Follow the directives for generating such a figure under the heading Graphs/Figures on this course's web site.

Restriction mapping of a yeast gene – Southern analysis

Below is shown an agarose gel electrophoresis of digests of yeasts *Saccharomyces cerevisiae* and *Saccharomyces pombe* genomic digests as you performed during the lab (Panel A). Following the migration, the gel was transferred and probed with part of a *Saccharomyces cerevisiae* gene (Panel B). (This file can be obtained from the course's web site under the heading Data>Southern)



6. How many times do the enzymes *Bam*HI and *Eco*RI cut the *S. cerevisiae* genome within the region spanned by the probe? BamHI (1), EcoRI (0)
7. Draw a possible restriction map of this genomic region of *S. cerevisiae*. Indicate on your map the potential region spanned by the probe. Your figure must be to scale. Follow the directives for generating such a figure under the heading Graphs/Figures on this course's web site.
8. What must have been the approximate minimum size of the probe?
9. The same gene can also be found in the genome of the closely related yeast, *S. pombe*. How many copies of the gene are there in the *S. pombe* genome? 5
10. The lane representing hybridization to the *S. pombe* genome shows two bands which are distinctly fainter. Give a possible explanation for what these hybrids may represent. Partial digest

Part III: Bioinformatics 2

1. Provide theoretical restriction maps of the 5 unknown genes available on this course's Web page. Indicate below each map the name of the gene and list the enzymes which do not cut.

   Give points if done and includes required info.

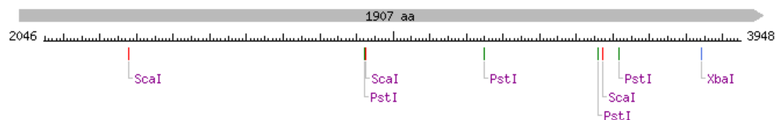
2. Compare the theoretical maps generated in the previous question to the experimental map of the unknown insert you analyzed in the lab ("**Restriction mapping of a recombinant plasmid (Pg. 25-31)**"). The unknown insert corresponds most closely to which gene? Cox

3. Indicate how many times each of the following enzymes cut within the unknown insert identified in question 2: *Acc*I, *Bgl*II, *Mbo*I, *Nco*I, and *Not*I. (3, 0, 17, 2 and 0 respectively)

4. Amongst the enzymes indicated in the previous question, which one cuts the most often within the DNA insert? Give a reason which would explain why this enzyme cuts more often than the others. MboI, shorter recognition sequence therefore higher probability

5. Provide a restriction map showing the positions of the restriction sites *Pst*I, *Sca*I, *Nco*I, and *Xba*I within the region spanned by positions 2046-3948 of the sequence "**Unknown sequence**" available on this course's Web page. Indicate below the map the definition of the gene and the enzymes which do not cut.

   Sequence digested with: NcoI, PstI, ScaI, XbaI



   NcoI does not cut. Gene name: Human mRNA for LCA-homolog.

6. The genome from which the sequence described in the previous question was obtained was digested with *Nco*I + *Sca*I and then migrated on an agarose gel. The gel was then transferred and probed with a sequence spanning the region 1491 – 2491 of the unknown sequence. How many bands and of what size would you expect on the Southern transfer? 2 bands: 645 + 1843bases.

Part IV: Lab performance

1. Submit a figure representing the agarose gel electrophoresis showing your plasmid isolation and the digests of the unknown recombinant plasmid. Make sure to include an appropriate legend. Follow the directives for figures on the Web page of this course and to include all the required information in the legend for the understanding and interpretation of the figure.

   (One point/criteria)
   Figure provided with Accompanying legend
   Figure title provided
   Lanes are labelled and easy to understand
   First lane is mw ladder
   Second lane is undigested control

MW sizes of standard are indicated
Loading is consistent between lanes
Digests are complete for the most part
Gel clearly shows bands and migration is satisfactory
Legend indicates parameters of migration: Voltage, agarose concentration, samples loaded

2. Provide the following information for your yeast genomic DNA isolation: ABS_{260}, ABS_{280}, Ratio ABS_{260}/ ABS_{280}, Concentration in µg/µL of undiluted preparation, Total yield in µg.

Assignment #3

Note: Part I of the assignment is to be submitted through Blackboard (See instructions on Blackboard). Parts II and III are to be submitted as a hard copy during your lab session.

Part I: Cloning and transformations

Sickle-cell anemia is caused by a mutation in the human β-globin gene. The three possible genotypes are homozygous for normal β-globin, heterozygous carrier (having both the normal and sickle - cell alleles), and homozygous for sickle cell anemia. Recombinant DNA technology has been used as the basis for prenatal diagnosis of sickle cell anemia. In a very high percentage of the cases observed, the normal human β-globin gene is carried on a 7600 bp human DNA fragment from a *Hpa*I digest, while the sickle-cell gene is carried on a 13000 bp *Hpa*I fragment.

Available to you are:

- A labeled sample of recombinant DNA consisting of a bacterial plasmid vector (4000 bp) carrying the 7600 bp *Hpa*I fragment from the normal human genome.
- A sample of genomic DNA from each member of a couple thought to be carriers of the sickle-cell trait and expecting their first child.
- A sample of the genomic DNA obtained from the fetal cells in the amniotic fluid from the uterus of the pregnant woman.

1. You perform a Southern on *Hpa*I digested DNA using the recombinant plasmid as a probe. What sized band (s) would you expect in each of the following cases?

   - Heterozygous carrier mother
   - Homozygous normal mother
   - Homozygous for sickle cell anemia fetus
   - Combined genomic DNA from a heterozygous carrier mother + homozygous normal father

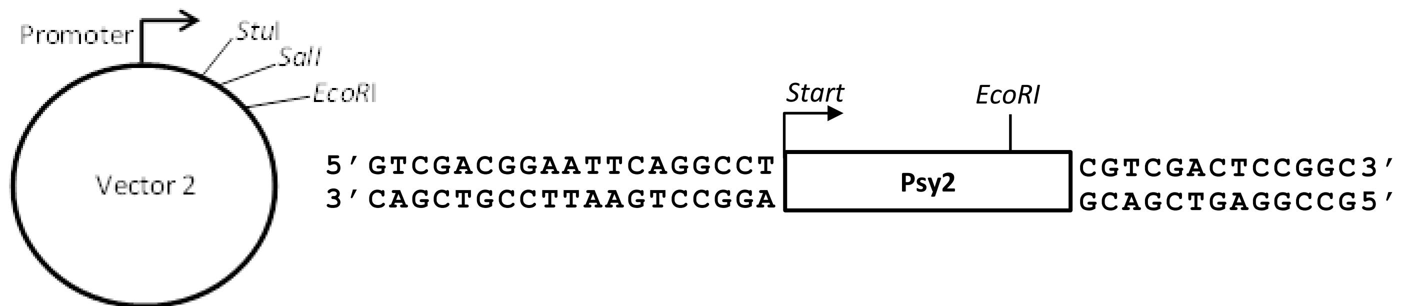
2. You have some DNA sequence from the psy2 gene of yeast as shown below. You decide to use PCR to amplify the psy2 sequence based on the flanking sequences shown below.

Indicate the set (s) of primer (s) you could use to amplify the entire psy2 coding sequence.

Set 1:	Set 2:	Set 3:
5' AGGCCG 3'	5' TCCGGC 3'	5' TCCGGC 3'
5' ACCGGG 3'	5' GCCGGA 3'	5' GCCGGA 3'

5' TCCGGCGGAATTCCAAGGCCT [psy2] CGTCGACTCCGGC 3'
3' AGGCCGCCTTAAGGTTCCGGA [psy2] GCAGCTGAGGCCG 5'

Consider the following information to answer questions 3-5:
You successfully amplify the psy2 sequence and wish to clone the PCR fragment into vector 2 in order to express it in yeast. The cloning sites available on this vector are shown below.



*Stu*I: 5'- AGG/CCT-3'   *Sal*I: 5'-G/TCGAC-3'   *Eco*RI: 5'-G/AATTC-3'   *Bam*HI: G/GATCC

There are two different ways to insert the amplified psy2 sequence into vector 2. Give the restriction enzyme(s) that you could use to cut the vector and the psy2 coding sequence for each of these.

3. **Directional cloning:**
   Cut vector and psy2 with:

4. **Non directional cloning:**
   Cut vector and psy2 with:

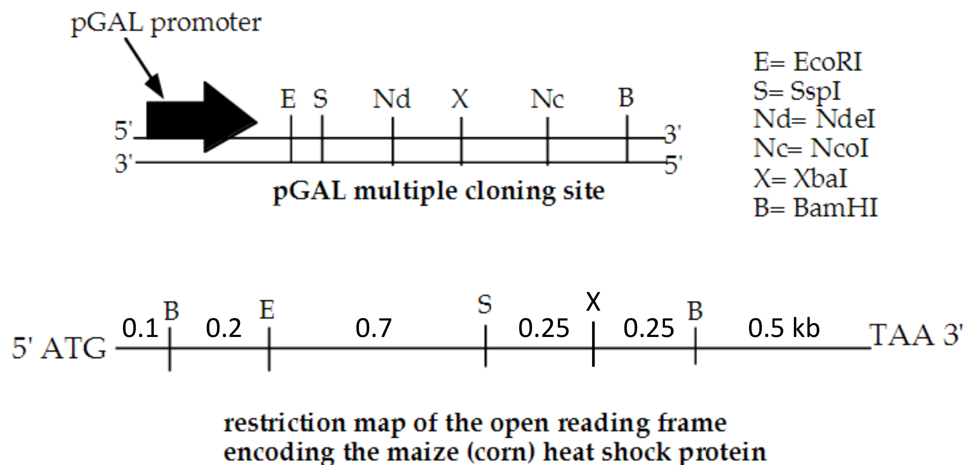
5. Which enzyme (s) could you use to verify the presence and orientation of the insert?

Consider the following information to answer questions 6-9:
Several lab groups independently carried out similar ligation/transformation steps that you performed in the lab to introduce GFP into the MCS of the phosphatase treated digested pUC vector. Both the vector and GFP sequence had been digested with the same single enzyme. The results obtained by each of the groups were as follows:

	# of colonies on LB-Amp plates			
Transformation into XL-1	Group 1	Group 2	Group 3	Group 4
a) Cut pUC + GFP insert	0	529	2	1125
b) Cut pUC	0	2	5	930
c) GFP alone	0	0	0	890
e) 10µl uncut pUC (0.5 ng/µL)	0	975	975	975
f) No DNA	0	0	0	925

6. Which group obtained the expected/desired pattern of results?
7. Which group would be expected to give the greatest percentage of blue colonies on X-Gal plates with the transformation mixture "a"?
8. Which transformation mixture would be expected to have the highest percentage of intramolecular ligation events?
9. What was the transformation efficiency of the competent XL-1 cells? Indicate the number of transformants expected per microgram of undigested DNA.

Consider the following information to answer questions 10-14:
You are interested in the directional cloning of the maize heat shock protein gene (2 kb) into the vector pGAL. You need to choose the restriction sites that you will include in your PCR primers and that you will use to cut the vector pGAL (5 kb). The following are maps of the multiple cloning site of pGAL and the DNA sequence that encodes the maize heat shock protein.



pGAL multiple cloning site

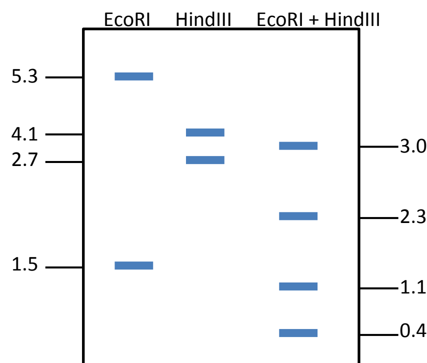
restriction map of the open reading frame encoding the maize (corn) heat shock protein

10. Which restriction site would you add to your forward primer?
11. Which restriction site would you add to your reverse primer?
12. You isolate plasmid DNA from a colony which you think has the insert of interest. To verify your assumption, you digest the recombinant plasmid with *Xba*I. What fragment size (s) is (are) expected?

To verify the orientation of the insert within the pGAL vector, you perform a digest on the positive recombinant with *Sal*I. What fragment sizes would you expect in each of the two orientations?

13. **Correct orientation:**
14. **Incorrect orientation:**

Consider the following information to answer questions 15-18:
You cloned an *Eco*RI digested DNA fragment into a unique *Eco*RI site of a vector. You identify a recombinant vector that you believe has the DNA of interest. To generate a restriction map of the recombinant plasmid, you take three individual samples of the plasmid and digest them with *Eco*RI, *Hind*III, and with both *Eco*RI and *Hind*III. You then run the digested DNA on an agarose gel to see the fragments. The gel is then subjected to a Southern hybridization using the *Eco*RI insert as a probe. Assume that the insert fragment is smaller than the vector.



15. Which fragment (s) on the gel would hybridize to the probe which represents the inert?
16. Which fragment (s) on the gel would hybridize to a probe representing the 5.3 kb band observed in the *Eco*RI lane?
17. Which fragment (s) in the *Eco*RI lane would hybridize to a probe representing the 4.1 kb band observed in the *Hind*III lane?
18. The cloning strategy used being non-directional, the recombinant plasmid may represent one of two possible orientations of the insert. You perform a hybridization with the 1.5 kb band observed in the *EcoR*I lane. Which fragment (s) would not be observed in the opposite orientation of that represented on the gel?

Consider the following information to answer questions 19-22:
You designed an experiment to clone a 0.8Kb fragment digested with *Eco*RI and *Hind*III into the unique *Eco*RI and *Hind*III restriction sites of a pUC vector. A double digest of the vector generates two fragments of 7.2 Kb and 0.016 kb. In order to minimize reassembly of the vector, the 0.016Kb fragment was removed from the double digest preparation of the vector. Transformations of the following samples were then performed:

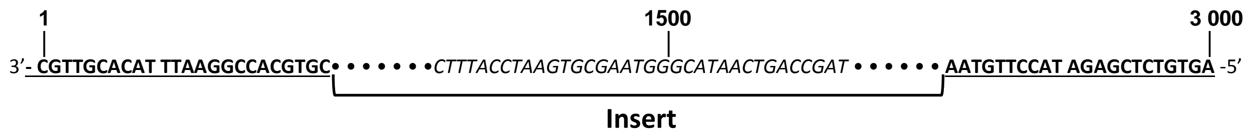
a. 30 ng of *Eco*RI digested vector which was treated with DNA ligase
b. 30 ng of *Eco*RI + *Hind*III digested vector which was treated with DNA ligase
c. 30 ng of *Eco*RI + *Hind*III digested vector mixed with 20 ng of the digested DNA fragment and then treated with DNA ligase

Below is indicated the number of colonies obtained for the different transformations. Unfortunately, the plates representing the different transformations were not labelled.

Plate 1: 5 colonies of color x + 40 colonies of color y
Plate 2: 348 colonies of color x + 2 colonies of color y
Plate 3: no colonies observed

19. What was the ratio between insert : vector in reaction "c"?
20. Which plate has the highest probability of containing colonies representing plasmid recombinants?
21. Which plate has the highest percentage of intramolecular ligation events?
22. What is the highest number of blue colonies observed?

You cloned an insert of approximately 3 kb into a single restriction site within the multiple cloning site of a vector. The partial flanking sequences of the multiple cloning site (indicated in **bold and underlined**) and of the insert sequence in the desired orientation (in *italics*) are indicated below. You wish to use colony PCR on various clones to identify those which possess the insert in the correct orientation (as indicated below).

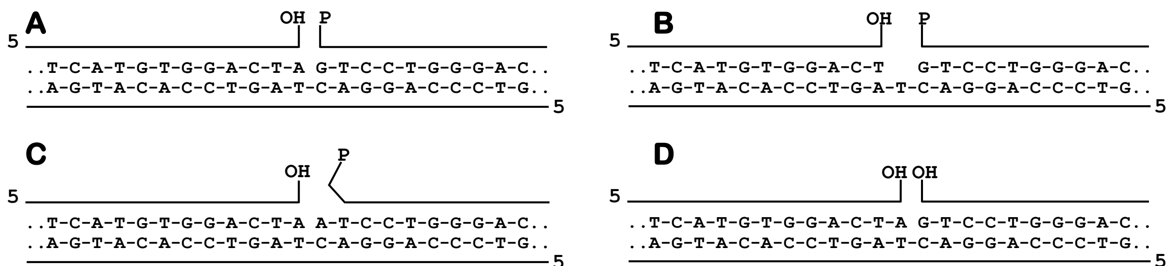


23. Which two primers could you use to identify clones which possess recombinant plasmids with the insert in the correct orientation?

Primers: (All sequences are written 5' to 3')

A. CGTTGCACAT
B. AGCTCTGTGA
C. AAATGGATTC
D. ATGTGCAACG
E. AGTGTCTCGA
F. TATGCCCATT
G. GCAACGTGTA
H. TCACAGAGCT

24. You've discovered a new virus which has a 10 kb double stranded DNA genome consisting of 65% A/T. How many times would you expect the enzyme *Hae*III (GG/CC) to cut this genome? Round off to the nearest whole integer.
25. Below are diagrams of double stranded DNA molecules some of which have mis-matches ("P" represents a 5' phosphate group) and single strand breaks. Indicate which single strand breaks could efficiently be closed by DNA.



Part II: Lab performance (10 points/question)

1. Ask a teaching assistant to verify that you can find your preparation of the pUC19 vector prepared in Lab exercise 2. Grant points to everyone. I've checked this.
2. Submit a figure representing your PCR of the GFP gene. Include an appropriate legend which includes the size of the amplicon observed. Give points if they obtained a product.
3. Submit a table providing the following information for each of the ligation mixtures of the different GFP sequences obtained with the three different forward primers as well as the ligation performed without an insert. Indicate the number of transformants obtained as number of colonies /mL. Give points if submitted with required info.
   - Reverse primer used for the amplification and mutagenesis
   - Total number of colonies observed
   - Number of white colonies
   - Number of blue colonies
   - Number of green fluorescent colonies
4. What was the transformation efficiency of the competent XL-1 cells provided? Show your calculation. Give points to everyone independently of what is submitted or even if nothing was submitted.
5. Submit a figure of the digestions performed on the amplicons obtained by colony PCR. Include an appropriate legend which indicates the sizes observed before and after the *Bam*HI digests. Give points based on quality of gel, they obtained a product and digestion worked.

Part III: Bioinformatics 3

1.
Step 1: Obtain the reverse sequence of the sequence with the accession number GBYX01460764.
Step 2: Obtain the complement sequence of the sequence from in step 1.
Step 3: Obtain the reverse complement sequence of the sequence from step 2.
Step 4: Obtain the reverse sequence of the sequence from step 3.
Step 5: Obtain the complement sequence of the sequence from step 4.
Indicate the first 20 bases of the final sequence obtained. Indicate the 5' and 3' ends.

AAAAAAAAAAAAAAAAAAAC

2. What would be the complement of the reverse sequence of the following sequence? Indicate the 5' and 3' ends.

5'-GATCGGATCCCATCTTATC-3'

3'-GATAAGATGGGATCCGATC-5'

3. The following primers were used in exercise 3 to amplify and mutagenize the GFP gene using pGFPuv as a template (the pGFPuv sequence can be obtained on this course's Web site):

   Primer 1: CGCCAAGCTTGCATGCCTGCAGGTCG
   Primer 2 CCGTCTCCGGGAGCTGCATGTGTCAG
   Primer 3: CGCCAAGCTTGCCATGCCTGCAGGTCG
   Primer 4: CGCCAAGCTTGACATGCCTGCAGGTCG

Submit the following information for each of the primers:

Are they "+" or "-"
Positions on the template of the 5' and 3' ends
Aligned primer sequences to the template indicating any mismatches

Ex.
```
    (1) THATSEQUENCE (12) (+ Primer)
        ||  |||||||
   (56) THISSEQUENCI (67) (Template)
```

or

Ex.
```
    (1) THATSEQUENCE (12) (+ Primer)
        **  *******
   (56) THISSEQUENCI (67) (Template)
```

```
230   CGCCAAGCTTGCATGCCTGCAGGTCG   255 (Template)
      ||||||||||||||||||||||||||
      CGCCAAGCTTGCATGCCTGCAGGTCG   26  (+primer)

1130   CTGACACATGCAGCTCCCGGAGACGG   1155 (Template)
       ||||||||||||||||||||||||||
26     CTGACACATGCAGCTCCCGGAGACGG   1  (-primer)

230   CGCCAAGCTTGC-ATGCCTGCAGGTCG   255 (Template)
      |||||||||||| ||||||||||||||
1     CGCCAAGCTTGCCATGCCTGCAGGTCG   27  (+primer)

230   CGCCAAGCTTG-CATGCCTGCAGGTCG   255 (Template)
      ||||||||||| |||||||||||||||
1     CGCCAAGCTTGACATGCCTGCAGGTCG   27  (+primer)
```

4. What is the expected size of the GFP amplicon? Approx. 925 bases

Consider the following information to answer questions 5-8:
In exercise 4, colony PCR will be used to screen potential GFP amplicons using the following primers:

Forward: AGCTCACTCATTAGGCACCCCAGGC
Reverse: ATCGGTGCGGGCCTCTTCGC

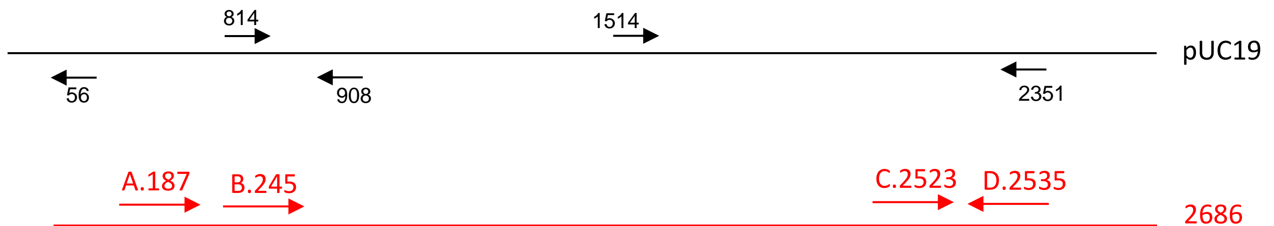
Initially determine whether these primers are directed against the pUC vector sequence, the GFP sequence or a combination of both.

5. What is the predicted size of the amplicon, if the chosen recombinant has the desired insert? Approx. 1063 bases
6. What is the predicted size of the amplicon, if the chosen recombinant does not possess an insert? 290 bases.
7. The colony PCR products will then be digested with *Bam*HI. Bands representing what sized fragments are expected in the case where the colony PCR product was obtained from a recombinant which has the desired insert? approx. 544, 343, and 176 bases.
8. The colony PCR products will then be digested with *Bam*HI. Bands representing what sized fragments are expected in the case where the colony PCR product was obtained from a recombinant which does not have an insert? Approx. 290

9. Map the alignment positions of each of the following primer sequences on the sequence of the pUC19 sequence. You may obtain the pUC19 sequence on the course's Web site.

   A. TGCGGTGTGAAATACCCT
   B. GCCATTCAGGCTGCGCAA
   C. GGGTTATTGTCTCATGAG
   D. GAGACAATAACCCTGATA

   Use a diagram, as shown below, to indicate the annealing positions of each of the primers. Each primer should be indicated by an arrow; where the head of the arrow represents the 3' end and the tail represents the 5' end. The direction of the arrow should indicate whether the primer is a "**Forward**" (→) or a "**Reverse**" (←) primer. The numbers associated with the arrows should represent the positions on the template corresponding to the 5' end of the primer.



10. Which of the primers indicated in the previous question, if any, would give an amplification product of at least 200 bp?

B and D give a product of at least 200bp

Assignment #4

Note: Part I of the assignment is to be submitted through Blackboard (See instructions on Blackboard). Parts II, III and IV are to be submitted as a hard copy during your lab session.

Part I

1. You have been asked to PCR amplify a specific sequence from cDNA that was synthesized from mRNA isolated from brain tissue. After you run your potential PCR product on an agarose gel containing ethidium bromide, you observe no bands when you visualize the gel using ultraviolet light. Why might this be the case? Choose all possible answers.
   A. The gene you are interested in is not expressed in brain tissue.
   B. An oligo dT was used instead of a oligo dA to prime the first strand cDNA synthesis.
   C. A specific primer whose sequence was that of the non-coding strand of the gene was used to prime the first strand cDNA synthesis.
   D. You used reverse transcriptase instead of Taq polymerase to synthesize the first strand of cDNA.

2. Starting with a single molecule of an mRNA template and sufficient primers, enzyme and all other co-factors for successful gene specific RT-PCR, how may PCR cycles are required to have 8 molecules of products with ends defined by both primers? By this I mean that you count only double stranded molecules that begin and end at the primer binding sites but lack any other sequence.

3. One strand of the extremely tiny gene Liliputian is shown below, with its start and stop codons underlined. Is this the template or the non-template strand?

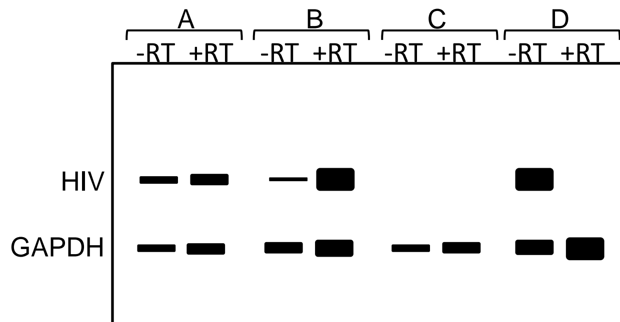
   5'-TGAGGCATCATCGGT<u>ATG</u>GCACCCTTAATGGGCATTGCACCCATAGTACGA<u>TAA</u>GCATGTCCTGAA-3'

4. I want to use RT-PCR to make a copy of the <u>entire</u> Liliputian gene. Indicate, the sequence of a primer of 6 nucleotides that could be used for the first strand synthesis of cDNA.

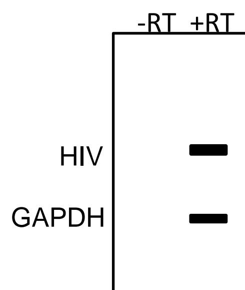
Consider the following information to answer questions 5-8.
The Liliputian gene is known to be repressed 5 fold when cells are grown in glucose as compared to growth in glycerol. To study the expression of this gene, a northern analysis was performed for a wild type strain as well as various mutants. Blots were simultaneously probed for the gene shown above, as well as a house keeping gene; GAPDH. A densitometric analysis was then performed. The densitometric values obtained for a wild type strain of the organism grown in glucose were 500 for both the Liliputian mRNA and the GAPDH mRNA. Indicate the expected densitometric value for the Liliputian mRNA in each of the following scenarios:

5. Wild type grown in glycerol (GAPDH value = 100)
6. Mutant strain with a mutation which enhances the promoter's activity 2 fold grown in glycerol (GAPDH value = 500)
7. Mutant strain with a mutation which changes the ATG codon to CTG grown in glycerol (GAPDH value = 250)
8. Mutant strain with a mutation which represses the promoter's activity 5 fold grown in glucose (GAPDH value = 500)

Consider the following information to answer questions 9-13.
RT-PCR is a commonly used diagnostic method used for the detection of HIV infections, caused by an RNA virus. Briefly, following infection, the viral RNA genome is converted to a double stranded DNA which integrates into the genome. Following integration the virus may remain dormant for several years. Following activation, the viral DNA is transcribed and initiates its replicative cycle; spreading from cell to cell. The following figure illustrates RT-PCR reactions done on the blood from different individuals. Reactions were carried out in the presence and absence of reverse transcriptase in the reaction mixture (+RT and –RT respectively). Simultaneous RT-reactions were performed on the house keeping gene; GAPDH.



9. Which individual (s) is (are) HIV positive?
10. Which individual (s) is (are) HIV Negative?
11. Which individual (s) is (are) actively replicating HIV?
12. Which individual (s) has (have) more copies of the HIV genome integrated in their genome?
13. The above experiment was repeated as previously, but on DNAse treated samples. One of the patterns observed is shown below. To which individual (s) could this pattern correspond to?



14. Which polymerase could read the following nucleic acid template and synthesize a new complementary strand? Choose all that apply.

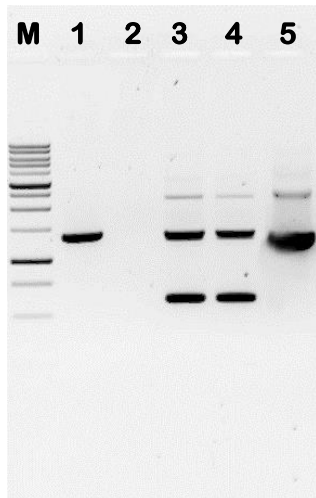
5’- AUUCUGGCUCGAAUGACUACUGGACC-3’

A. A DNA dependent DNA polymerase
B. A DNA dependent RNA polymerase
C. An RNA dependent RNA polymerase
D. An RNA dependent DNA polymerase

15. Indicate the sequence in the 5’ to 3’ direction of the first 6 bases which would be synthesized by the chosen polymerase.

Part II: Lab Exercises

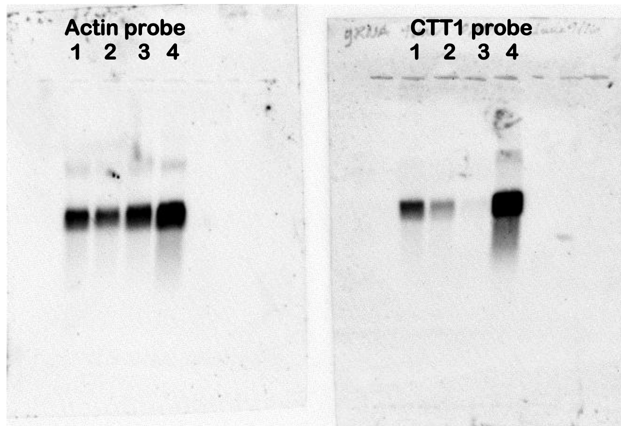
Below is a gel of RT-PCR reactions similar to the ones you performed in exercise 5. (You may obtain a copy of these results on this course’s Web page under the heading data>rtpcr)



Reaction	Template
PCR-1	**R** (Step I)
PCR-2	**DR** (step II)
PCR-3	**RT-R** (step III)
PCR-4	**RT-DR** (step III)
PCR-5	Yeast genomic DNA

1. What is the purpose of the DNAse treatment?
2. What is the purpose of the PCR reaction using RNA NOT treated with either DNAse or RT?
3. What were the sizes of the RT dependent and RT independent PCR products?
4. What does the difference in size between the PCR products tell you about the gene from which this RNA is derived from?

Below is an autoradiogram obtained from a northern hybridization of yeast total RNA using the CTT1 gene and the actin gene as probes. (You may obtain a copy of these results on this course's Web page under the heading data>northern)



1. **Cultures exposed to 0.85M NaCl for 0 minutes.**
2. **Cultures exposed to 0.85M NaCl for 30 minutes.**
3. **Cultures exposed to 0.85M NaCl for 60 minutes.**
4. **Cultures exposed to 0.85M NaCl for 120 minutes.**

5. Submit a table of the densitometric analysis of this northern hybridization. Your table must include the raw data (the values for each of the areas), the normalized values (CTT1 reading/actin reading) for each of the growth conditions and the relative expression as compared to growth in the absence of osmotic shock.

Part III: Lab performance

1. Indicate the ABS_{260}, ABS_{280} and the ABS_{260}/ABS_{280} ratio of your RNA preparation (Pg. 47).
2. Submit a figure of the RNA gel generated in lab exercise 5 (Pg. 48). Include an appropriate figure legend.
3. Submit a figure of the RT-PCR gel generated in lab exercise 5 (Pg. 50-52). Include an appropriate figure legend, which indicates the sizes of the products observed in each lane.
4. Submit a figure of the RT-PCR gels (CTT1 and actin) generated in lab exercise 6 (Pg. 54-56). Include an appropriate figure legend, which indicates the sizes of the products observed in each lane.

Part IV: Bioinformatics 4 & 5

1. What is accession number and source organism of the first nucleotide record for the hemoglobin, alpha 2 mRNA from *Bos taurus*? NM_001077422.3
2. What is accession number and source organism of the first nucleotide record for the hemoglobin, alpha A mRNA from *Gallus gallus* (chicken)? NM_001004376.3
3. What is accession number and source organism of the first nucleotide record for the human fetal hemoglobin? NM_000559.2

Questions 4-7 Give points if done

4. What is the percentage of identity at the nucleotide level between the hemoglobin genes from each of the following pairs of organisms:

    A. Human alpha hemoglobin to cow hemoglobin
    B. Human alpha hemoglobin to chicken hemoglobin
    C. Cow hemoglobin to chicken hemoglobin
    D. Human alpha hemoglobin to human fetal hemoglobin

5. Which pair of sequences, if any, are analogues?
6. Which pair of sequences, if any, are paralogues?
7. What is the percentage of identity at the protein level between the hemoglobin proteins from each of the following pairs of organisms:

    A. Human alpha hemoglobin to cow hemoglobin
    B. Human alpha hemoglobin to chicken hemoglobin
    C. Cow hemoglobin to chicken hemoglobin
    D. Human alpha hemoglobin to human fetal hemoglobin

8. What is the accession number for the alcohol dehydrogenase protein record of *Bos taurus* obtained through Blink? XP_010826388
9. What is the accession number for the alcohol dehydrogenase protein record of *Mus musculus* obtained through Blink? BAE43215
10. What is the accession number for the alcohol dehydrogenase protein record of *Saccharomyces cerevisiae* obtained through the general search? AAA34411.1
11. Which pair of amino acid sequences, if any, are analogues? Give points if done
12. Which pair of amino acid sequences, if any, are orthologues? Give points if done
13. What is the potential function of the unknown protein from *Danio rerio*? cytochrome c oxidase subunit I;
14. The unknown protein from *Danio rerio* shares the highest level of identity with a protein from which organism? Homo sapiens
15. What are the definition corresponding to the viral 1 & 2 sequences? Influenza A virus (A/New York/492/2003(H1N2)) segment 4, complete sequence & Influenza A virus (A/New York/492/2003(H1N2)) segment 6, complete sequence.
16. From which organism do the viral 1 & 2 sequences come from? Influenza A virus (A/New York/492/2003(H1N2)) for both
17. What is the name of the protein products from the longest ORFs from the viral 1 & 2 sequences? hemagglutinin & neuraminidase

18. What are the gene names which code for the longest ORFs from the viral 1 & 2 sequences? HA & NA
19. In which reading frame were the longest ORF from the viral 1 & 2 sequences found? -1
20. Was the longest ORF found for the viral 1 sequence on the inputted sequence or its reverse complement? Reverse complement
21. What is the percentage identity between the translated ORF from viral 1 and the most closely related protein?98%
22. When comparing the longest translated ORF from the viral 1 sequence to the most closely related protein, what percentage of the amino acids represent conservative substitutions? all
23. What snp has the viral 3 sequence acquired? Indicate the position and base change. Ex. C118 to A. Indel –C at position 1537 and + A at position 1600
24. What type of amino acid change, if any, does this snp lead to?

Assignment #5

Note: Part I of the assignment is to be submitted through Blackboard (See instructions on Blackboard). Parts II, III and IV are to be submitted as a hard copy during your lab session.

Part I

Refer to the following description to answer questions 1-6

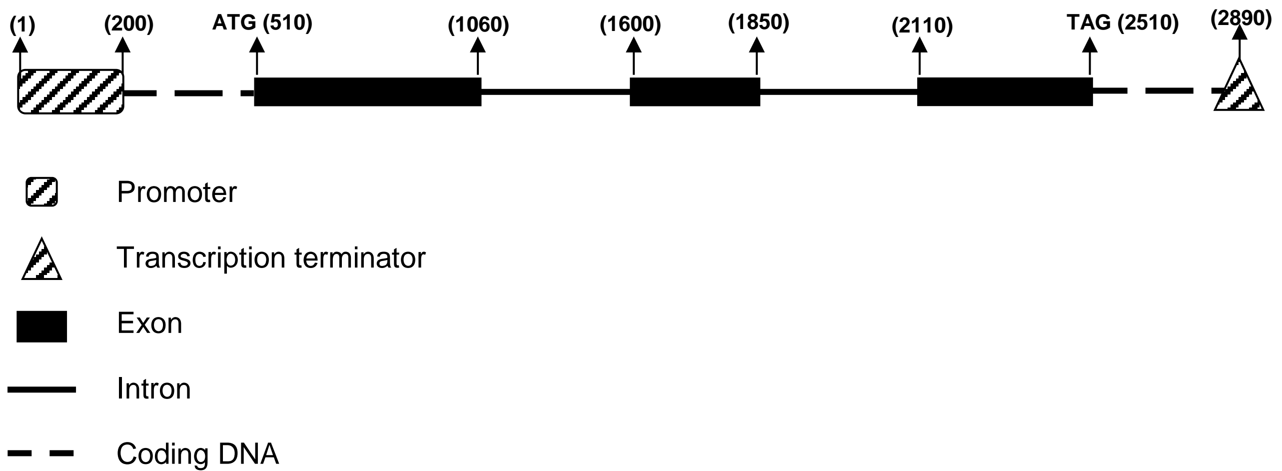
Below are 210 consecutive base pairs of DNA that includes only the beginning of the sequence of gene X. The underlined sequence (from position 20-54) represents the promoter for gene X and the underlined and italicized sequence (from position 71-90) encodes the gene X ribosome binding (RBS) site. Transcription begins at and includes the T/A base pair at position 60 (underlined).

```
   1        10        20        30        40        50        60        70
   I--------I---------I---------I---------I---------I---------I---------I
5' ATCGGTCTCGGCTACTACATAAACGCGCGCATATATCGATATCTAGCTAGCTATCGGTCTAGGCTACTAC
3' TAGCCAGAGCCGATGATGTATTTGCGCGCGTATATAGCTATAGATCGATCGATAGCCAGATCCGATGATG
                                         Promoter
            80        90        100       110       120       130       140
            I---------I---------I---------I---------I---------I---------I
5' CAGGTATCGGTCTGATCTAGCTAGATGCTCTTCTCTCGAACCCGCGGGGGCTGTACTATCATGCGTCG
3' GTCCATAGCCAGACTAGATCGATCTACGAGAAGAGAGCTTGGGCGCCCCCGACATGATAGTACGCAGC
       RBS
           150       160       170       180       190       200       210
   ---------I---------I---------I---------I---------I---------I---------I
5' TCTCGGCTACTACGTAAACGCGCGCATATATCGATATCTAGCTAGCTATCGGTCTCGGCTACTACGTAAA
3' AGAGCCGATGATGCATTTGCGCGCGTATATAGCTATAGATCGATCGATAGCCAGAGCCGATGATGCATTT
```

1. What are the first 6 nucleotides of the mRNA from gene X? UAGGCT
2. What are the first 4 amino acids encoded by gene X? Use the one letter amino acid code to indicate your answer. Ex. TRCG MLFS
3. You have found a mutant of gene X. The mutation represents an SNP which changes the G/C base pair at position 110 (underlined) to T/A. Would the mRNA expressed from this version of gene X be longer, shorter, or the same as that produced from the normal gene X? Same
4. If the mutant mRNA can be translated, would you expect the protein to be longer, shorter, or the same as that produced from the normal gene X? Shorter
5. You have found another mutant of gene X. The mutation represents an SNP which changes the G/C base pair at position 110 (underlined) to C/G. Do you expect that the protein produced will have a similar level of activity as the normal protein X? Similar
6. What is the length of the 5' UTR? 96b

Refer to the following description to answer questions 7-11

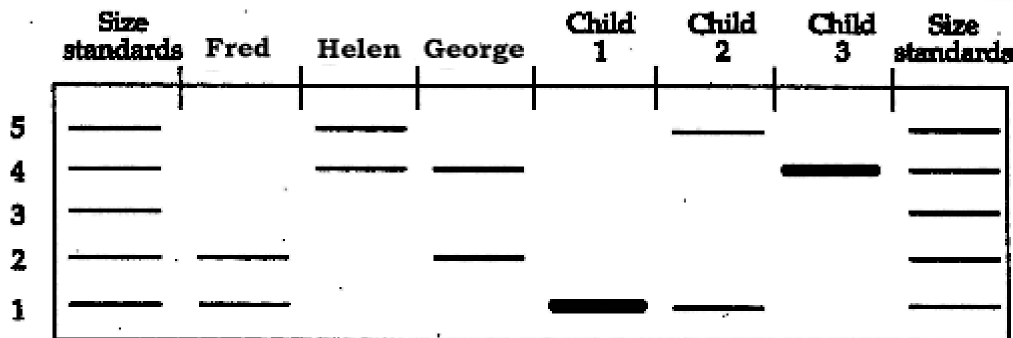
The following schematic represents a yeast gene and its various elements:



Indicate the predicted sizes of each of the following:

7. Pre-mRNA (non processed) 2690
8. mRNA 2090
9. Protein 401
10. mRNA from a gene with a point mutation creating a stop codon at position 1698 2090
11. Protein from a gene with a point mutation creating a stop codon at position 1698 215

12. Fred is married to Helen, who was previously married to George, now deceased. George and Helen conceived one child together and adopted one child. Fred and Helen have also conceived one child. All members of Helen's current family have had DNA fingerprinting done at a single VNTR locus. Unfortunately, the sheet that identified each child has been misplaced. Identify which fingerprint in each lane (in lanes 5, 6, and 7) correspond to each child.



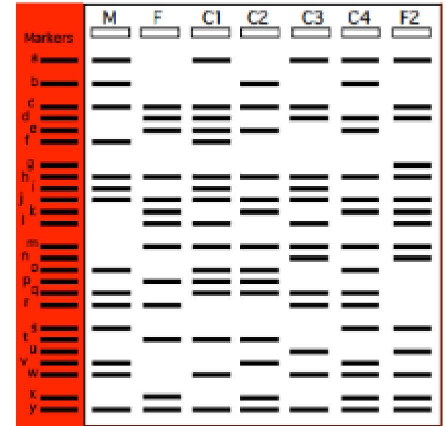
Fred and Helen's conceived child Child ___2_

George and Helen's conceived child Child ___3_

George and Helen's adopted child Child ___1_

Refer to the following description to answer questions 14-17

The following diagram shows a Southern blot of restriction digested genomic DNA from an elf Mother (M) and elf Father (F) and four potential elf children (C1 to C4) probed with a VNTR DNA sequence. The restriction enzyme used was *Not* I. Another elf (F2) is claiming to be the father of child C4. Assume the mother, M, is the real mother of these four children.



13. Which children are the biological children of the elf Father (F)? 1, 2 and 4
14. Which children if any, could be elf F2's children? C3
15. Consider the data in the lane labelled "markers". According to this lane only, what is the minimum number of loci which are being examined? 13
16. If the number of loci indicated in the previous question represents the total number of loci being examined, the profile of C4 must include a minimum of how many heterozygous loci? 3

Refer to the following description to answer questions 17-21

The following schematic shows the region of a gene associated with hairy toes (HTO). The hairy toes phenotype is due to recessive mutations which create an RFLP within a *Hae*III restriction site at position 550 (*Hae*III sites are indicated by arrows). Note that there may be other snp, but these are unrelated to the phenotype of hairy toes. This region of the genome was examined in different individuals by performing a Southern analysis on genomic DNA samples digested with *Hae*III and probed with a DNA fragment spanning the region from 425-650. The results of the hybridization are shown in the table below.

Individual	Sizes of bands observed (bp)
A	450 and 625
B	475, 300 and 150
C	300 and 150
D	475, 450 and 150

17. Which individual (s) have a genotype which would give them the hairy toes phenotype? C
18. Which individual (s) have a genotype which would give them the normal phenotype? A, B and D
19. Which individual (s) is (are) a carrier of the hairy toes genotype, but have a normal phenotype? D
20. How many different alleles of the region were detected among the four individuals? 4

21. Below are the sequences of portions of the yfg1 mRNA and three yfg1 mRNA variants (A, B, and C) from zebrafish embryos.

    5'...GAUGAAAGAUCAGGUCUGAAUGUAU...3' yfg1 mRNA

    5'...GAUGATTCAUCAGGUCUGAAUGUAU...3' Variant A

    5'...GAUGAAAGAUCAGGACAGAAUGUAU...3' Variant B

    5'...GAUGAAAGAUCAACUCUGAAUGUAU...3' Variant C

    You've created a probe which is 100% complementary to the sequence of the yfg1 mRNA: 5'...TTCAGACCTGATCTTTCATC...3'. You hybridize the probe to a blot of total zebra fish RNA at room temperature (20˚C), and then want to adjust the stringency by washes at 50˚C. Which of the four mRNAs will you "see" as bands on film when you develop your northern blot?

    A. yfg1 alone
    B. yfg1 and A
    C. yfg1 and B
    D. yfg1 and C
    E. all of them
    F. none of them

Indicate how each of the following conditions would affect the Tm of nucleic acid hybrids; increase, decrease or no effect.

22. Increased temperature No effect
23. Increased urea concentration Decrease
24. Higher G:C content Increase

25. You wish to use a probe of 250 bases which is 40% G/C to perform a hybridization at 40°C in a buffer containing 0.5 M NaCl. Assuming 100% complementarity, your hybridization buffer should contain what maximum percent concentration of formamide? Assume optimal hybridization will be at 5°C below the Tm. 67

Part II: Lab performance Give full points if done.

PCR amplification of *Apo*C2 VNTR (Lab exercise 7, pg. 61)

1. Submit a figure of the VNTR profiles of the class. Include an appropriate figure legend which includes a brief analysis. The analysis should include the sizes of the different alleles observed identified alphanumerically, the percentage of individuals examined possessing each of the different alleles, and whether each individual was homozygous or heterozygous.

PCR amplification of *Apo*B RFLP (Lab exercise 7, pg. 61)

2. Submit a figure of the RFLP profiles of the class. Include an appropriate figure legend which includes a brief analysis. The analysis should include the different alleles observed labelled alphanumerically, the percentage of individuals examined possessing each of the different alleles, and whether each individual was homozygous or heterozygous.

Stringency – DNA melting curves (Lab exercise 7, pg. 64)

3. Submit a graph of the DNA melting profile obtained under the assay condition assigned. The graph should represent the ABS260 Vs temperature. Include an appropriate figure legend which includes a brief analysis. The analysis should report the approximate Tm under the conditions assayed.
4. Submit a table presenting the different melting temperatures obtained under the different conditions assayed by the class.

Protein quantification – Bradford assay (Lab exercise 8, pg. 66)

5. Submit a graph of your protein standard curve generated from the Bradford assay. Include an appropriate figure legend which includes a brief analysis. The analysis should show the calculation of the protein concentration within the muscle tissue extract you prepared.

Protein gel electrophoresis (Lab exercise 8, pg. 67)

Part III: Bioinformatics

You should now be quite familiar with the NCBI site and be able to complete the following exercise with relatively few directives. Consider this as a practice run for the bioinformatics section on the final exam.

Use the following sequence to answer questions 1-10

```
TGAAGCCCTTGGATTCTTGAACGAGGATCACTGGATGGGGAGAGAGAACTCAGGAGGTGGTGTTGAAGGG
CTGGGATTACAAAGACTCGGATATGTCCTAGAAGAGATGAGTCGCATACCAGGAGGAAGGATGTATGCAG
ATGACACTGCTGGCTGGGACACCCGCATCAGCAGGTTTGATCTGGAGAATGAAGCTCTAATCACCAACCA
AATGGAGAAAGGGCACAGGGCCTTGGCATTGGCCATAATCAAGTACACATACCAAAACAAAGTGGTAAAG
GTCCTTAGACCAGCTGAAAAAGGGAAAACAGTYATGGACATTATTTCGAGACAAGACCAAAGGGGGAGCG
GACAAGTTGTCACTTACGCTCTTAACACATTTACCAACCTAGTGGTGCAACTCATTCGGAATATGGAGGC
TGAGGAAGTTCTAGAGATGCAAGACTTGTGGCTGCTGCGGAGGTCAGAGAAAGTGACCAACTGGTTGCAG
AGCAACGGATGGGATAGGCTCAAACGAATGGCAGTCAGTGGAGATGATTGCGTTGTGAAGCCAATTGATG
ATAGGTTTGCACATGCCCTCAGGTTCTTGAATGATATGGGAAAAGTTAGGAAGGACACACAAGAGTGGAA
ACCCTCAACTGGATGGGACAACTGGGAAGAAGTTCCGTTTTGCTCCCACCACTTCAACAAGC
```

1. What is the name of the organism from which this sequence was obtained? Zika virus
2. What is the probability that this gene is a true match? 100%
3. Is the source sequence RNA or DNA? RNA
4. What is the accession number of a corresponding sequence with 100% coverage? KY317937.1
5. What is the name of the protein product encoded by this sequence? Polyprotein
6. The protein product from this sequence is how many amino acids long? 3423 aa
7. What is the potential function of this protein? RNA-directed RNA polymerase
8. The above sequence was cloned into the *Pst*I site of pUC19. What fragment sizes would be expected following a digest with *Xba*I of recombinants containing the insert of interest in either orientation? 262 + 3116 and 430 + 2948
9. Which of the following primers could be used to perform a reverse transcriptase reaction on the coding RNA? accept either answer.

   A. GAGCAAAACGGAACTC
   B. AAGAAGTGGTGGGAGC
   C. TTTTGCTCCCACCACTG
   D. GCTTGTTGAAGTGGTG

10. What is the amino acid length of the longest ORF on the negative strand of the sequence obtained in question 4? 182 aa.

Refer to the record with the accession number NM_001190338 to answer questions 11-15

11. What is the accession number of the first human nucleotide homologue of this sequence? NM_004367.5
12. What type of homologue is the human nucleotide sequence? analogues
13. What is the percentage amino acid identity between the human protein homologue and the protein encoded by the initial sequence? 75.41%
14. What type of amino acid substitution is most abundant between the two homologues (human and original protein)? Non conserved.
15. Consider the following sequence: CGTAGCGACTTTTGGGTTCACAACCACCGTACTC. Does this sequence represent the original sequence, the reverse sequence, the complement sequence, or the reverse complement sequence? Reverse