Welcome to the First Genetics Exam! (The aftermath)

(Fall 2000)

Scoring

Your exam will have two numbers near the bottom left of your answer sheet. The first is the raw score, the sum of the number of points you got for each question. The highest number possible was 121, but we threw out your worst questions, up to a total of 10 points. The second (circled) number is the normalized score, calculated according to the formula that will appear in full below once we've graded all of the exams.

$$\text{normalized score} = 75 + 20 \cdot (\text{raw score} - \mathbf{K_1}) / \mathbf{K_2}$$

The purpose of this complicated formula is to put the scores in the range of 50 to 100, where 90-100 is A, 80-90 is B, etc. The factors were chosen so that the normalized scores correspond to our perception of what constitutes an A and so forth. In this way, you know how this and all of your exam grades contribute to your final grade. There was no curve in this exam nor will there be in any future exam or for the final grade.

The questions

3. (2) The term "DNA", often encountered in the scientific literature, can be explained as follows:

A. It is an abbreviation of deoxyribonucleic acid

B. It is an abbreviation of dihydroneuraminic aldehyde

C. It is an abbreviation of the National Dyslexic Association

D. Originally pronounced "deh-nah", the word is a contraction of "denarius", a Roman coin, drawing on the metaphor of genes as the coins of inheritance.

You're just lucky this is an open book exam.

5. (8) Conclude as much as you can from the following results of an experiment in which artificial RNA was translated in an in vitro translation system.

CUCUCUCU. . .	ArgCysArgCys. . .
CUUCUUCUU. . .	CysLysAlaCysLysAla. . .

The code is not a nonoverlapping duplet, triplet, or quadruplet code. It is neither a duplet overlapping code nor a triplet code with one-base overlaps. Hard to see how a variable code could work. Triplet with two-base overlaps works, as does quintuplet nonoverlapping and others. Presuming the code is triplet, overlapping, UCU must encode Cys and CUC must encode Arg.

Not nonoverlapping duplet: If it were, CU CU CU would encode a single-amino acid repeating peptide.

Not nonoverlapping triplet: If it were, CUU CUU CUU would encode a single amino acid repeating peptide.

Not nonoverlapping quadruplet: If it were, CUCU CUCU would encode a single amino acid repeating peptide.

Not overlapping duplet: If it were, CU and UC would encode Arg and Cys both would appear in the peptide encoded by CUUCUUCUU….

Not triplet with one-base overlaps: If it were, CUCUCU… would encode single amino acid repeating peptides.

UCU encodes Cys: UCU appears in CUCUCU… and CUUCUU…. Cys is encoded by both.

CUC encodes Arg: CUC appears in CUCUCU… but not CUUCUU…. Arg is encoded only by the first.

6. (15) A culture of the bacterium *Bacillus subtilis* was incubated with radioactive phosphate for twenty four hours, a long period of time, since the bacterium requires only 30 minutes to divide or (in a separate experiment) for twenty minutes. After the incubation, total DNA was isolated (and subjected to RNAse to destroy any RNA). At the same time, total RNA was isolated (and subjected to DNAse to destroy any DNA). The two samples (DNA and RNA) were hydrolyzed to produce their component bases, and the radioactive bases were analyzed, to give the results shown in Table 3.

Table 3: Base composition of DNA/RNA[a]

	G	A	T or U	C
DNA				
20 min	15.8%	34.1%	35.0%	15.1%
24 hrs	15.6%	34.2%	34.8%	15.4%
RNA				
20 min	16.7%	33.2%	32.7%	17.4%
24 hrs	29.5%	28.1%	20.4%	22.0%

[a]Fraction of radioactivity in each of four bases after incubation with radioactive phosphate for indicated interval

6a. Explain in fewer than 10 words why the radioactive base composition of RNA resulting from the 24-hour incubation is so different from that of DNA.

Most of the radioactive bases come from rRNA.

After 24 hrs of labeling (about 48 generations), all RNA is labeled equally, but about 80% of cellular RNA at any given moment is rRNA. So most of the radioactive bases come from rRNA, and rRNA happens to have something close to the base composition shown for RNA at 24 hrs. The genes that encode rRNA comprise only a tiny fraction of the total DNA, which happens to have a much different base composition.

6b. Explain in fewer than 20 words why the radioactive base composition of RNA resulting from the 20-minute incubation is so similar to that of DNA.

Most of the radioactive bases come from mRNA.

It is important to see that the relative amounts of rRNA and mRNA are the same at both times. And why not? The experiment is just taking a snapshot at different times. There's no reason to think that the bacteria is feeling any different after 20 minutes after labeling than after 24 minutes. What is different is that the radioactivity has more readily gotten into mRNA than rRNA. Why? mRNA is unstable, continuously being made and unmade, while rRNA is stable. If a revolutionary type of plastic were invented that

replaced all others, it would be incorporated into new refrigerators and new toothbrushes, but the fraction of toothbrushes with the new plastic in it would rise much faster. Toothbrushes turn are bought and tossed very fast, while it takes a long while for old refrigerators to retire.

The fraction of RNA in Okazaki fragments is very small. Only a tiny fraction of DNA is replicating at any given moment, and the RNA primers are rapidly replaced.

6c. Draw a picture of an autoradiogram of a gel with two lanes. The first lane contains RNA from the 24-hour incubation and the second contains RNA from the 20-minute incubation. An autoradiogram shows the position of radioactivity.

Shown to the right.

Radioactive RNA labeled for 24 hrs is dominated by rRNA (2.9 kb, 1.5 kb, and 0.1 kb) as well as tRNA (0.1 kb). Radioactive RNA labeled for 20 minutes is mostly mRNA. The rRNA is there, but obscured.

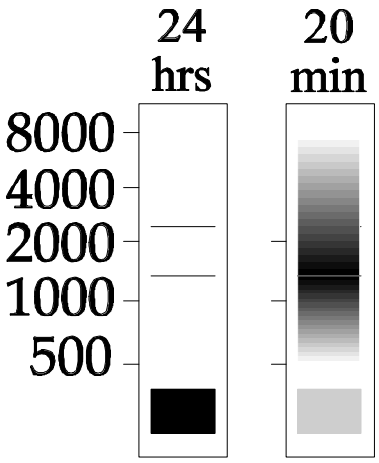
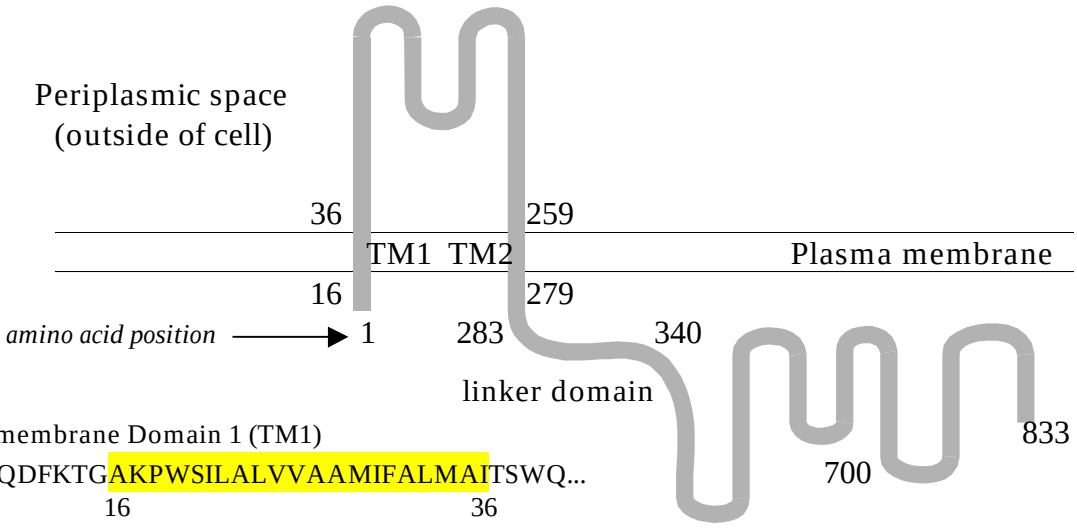


Fig. 1



Transmembrane Domain 1 (TM1)

MNGRYSPSRQDFKTGAKPWSILALVVAAMIFALMAITSWQ...

1 16 36

Transmembrane Domain 2 (TM2)

...QSVREQRARIFLGSVSVGLCIYIISLVYRLRRKTAWL...

259 279

7. (9 pts) Chang & Winans (1992), and Turk et al. (1994) made deletions in the *vir*A gene that removed coding region but maintained the same reading frame (see example below), transformed each mutated gene into a *A. tumefaciens* strain that lacked *vir*A, and then measured VIRA activity [none (-), very low (±), high (+), very high (++)].

a. In 10 words or less, where does AS bind to VIRA?

Somewhere between amino acids 283 and 304.

b. In 10 words or less, where does GLC bind to VIRA?

Somewhere between amino acids 40 and 240.

c. In 10 words or less, what part of VIRA most likely contains the active site for enzymatic activity?
Somewhere between amino acids 340 and 700.

8. (16 pts) Doty et al. (1996) randomly mutagenized the *vir* A gene with hydroxylamine to introduce single base substitutions. They introduced each mutant gene into a *A. tumefaciens* strain that lacks *vir*A, and then measured VIRA activity [none (-), very low (±), high (+), very high (++)].

Amino Acid Substitution in VIRA Protein Due to Mutation in *vir*A Gene	VIRA Activity: no activators	high AS	low AS + low GLU
no mutation (wt)	±	++	++
Ser20 to Phe	±	±	±
Arg88 to Trp	±	++	±
Asp139 to Asn	±	++	±
Arg209 to Cys	±	++	±
Glu210 to Lys	±	++	±
Gly268 to Asp	±	±	±
Thr284 to Met	±	±	±

8a. What kind of change is Arg88 to Trp? (silent, conservative, charged hydrophilic to uncharged hydrophilic, hydrophilic to hydrophobic, etc.)
large charged hydrophilic to large hydrophobic

8b. What kind of change is Glu210 to Lys?
large - charged hydrophilic to large + charged hydrophilic

8c. What kind of change is Gly268 to Asp?
small uncharged hydrophilic to large charged hydrophilic

8d. In 20 words or less, does the data from Doty et al. change your view of where GLC binds to VIRA? If so, how?
no, same part of VIRA is implicated in binding GLC

8e. In 20 words or less, does the data from Doty et al. change your view of where AS binds to VIRA? If so, how?
yes and no; the same area as before is implicated, but in addition so are transmembrane domains 1 and 2 implicated in binding AS

8f. Given the kind of change involved in Gly268 to Asp, come up with an alternative explanation for the effect of this mutation that does not involve direct binding to AS and / or GLC (30 words or less).
since Gly268 to Asp is a big change in transmembrane domain 2, maybe that stretch of mutant VIRA cannot cross the membrane properly. The rest of the protein would not be in the proper place and conformation, and maybe this disrupts its ability to be activated by AS.

9. (8 pts) Aspartate transcarbamoylase (ATCase) is a multimeric enzyme of identical subunits that catalyzes the first committed step in the biosynthesis of pyrimidine nucleotides. The *pyr*B gene encodes the monomer subunit of ATCase. Use the information given below to figure out the effects of two different mutations in *pyr*B on ATCase structure and function.

a. Mutation 1 was originally found in a strain of *E. coli* that completely lacked ATCase activity. The mutant *pyr*B- gene was cloned into a plasmid and then introduced into wildtype *E. coli*. In 30 words or less, deduce as much as you can from the table below about ATCase structure / function and the effect of mutation 1.

Strain	Amount of Subunit Protein	Quaternary Structure(s) Seen for Subunit Protein	ATCase Activity (Fraction of wt)
1. wildtype *pyr*B	60 units	hexamers only	1
2. *pyr*B- (mutation 1)	60	monomers only	0
3. wildtype pyrB on chromosome and pyrB- (mutation 1) on plasmid	60 (30 of each kind)	1/2 in monomers & 1/2 in hexamers	$^1/_2$

ATCase acts as a hexamer, or in other words the enzyme is composed of six identical subunits. The subunit encoded by *pyrB*- (mutation 1) cannot interact to form the hexamer quaternary structure needed for activity. Since only 1/2 of the subunits in the heterozygote are wildtype and can interact to form quaternary structure, only half as many hexamers can form and the activity is half of that normally seen.

b. Mutation 2 was originally found in a strain of *E. coli* that completely lacked ATCase activity. The mutant *pyr*B- gene was cloned into a plasmid and then introduced into wildtype *E. coli*. In 30 words or less, deduce as much as you can from the table below about ATCase structure / function and the effect of mutation 2.

Strain	Amount of Subunit Protein	Quaternary Structure(s) Seen for Subunit Protein	ATCase Activity (Fraction of wt)
1. wildtype *pyr*B	60 units	hexamers only	1
2. *pyr*B- (mutation 2)	60	hexamers only	0
3. wildtype pyrB on chromosome and pyrB- (mutation 2) on plasmid	60 (30 of each kind)	hexamers only	$^{57}/_{64}$

The subunit encoded by *pyrB*- (mutation 2) can still interact to form the hexamer quaternary structure, but something else is wrong. A good possibility is that the active site is somehow affected. With half good subunits and half bad subunits in the heterozygote, there are 64 different combinations of subunits in hexamers (2 possibilities at each position and six positions in hexamer; $2^6 = 64$). We are told that the heterozygote has 57/64 activity compared to normal. One combination, accounting for 1/64 of the the hexamers, is all bad subunits and clearly doesn't work (we were told that in line 2 of the table). We need six more combinations that don't work. The next simplest possibility are the six combinations where 5 subunits are bad and 1 subunit is good. If we assume that these combinations don't work, then all other combinations have to work to account for 57/64 activity. In other words, as long as 2 or more subunits in a hexamer are good, the hexamer works.

10. (5 pts) In a particular bacterial species, G+C = 4(A+T) and A+G = C+T.

a. What is the nucleotide composition of this organism?

b. In 10 words or less, what assumption did you make to reach your answer in part a.?

Obviously, we are talking about DNA here, so A + G + C + T = 100. Substituting the first piece of information leads to A + T + 4(A + T) = 100 or A + T = 20. Therefore, G + C = 80. The second piece of information sets certain limits on the possible solutions (for example, G = 70, C = 10, A = 7, T = 3 agrees with the first piece of information, but does not agree with the second). However, even within the limits set by the second piece of information, there are many possible solutions.

If you assume that the DNA is double stranded, then A = T and G + C. So the problem resolves itself into A = 10, T = 10, C = 40, and G = 40, which satisfies all criteria.

If you assume that the DNA is single stranded (although no cellular life forms have ss DNA genomes), then you can come up with several alternatives. Here are three of them.

A	T	C	G	G+C=4(A+T)		A+G=C+T	
0	20	30	50	80	4(20)	50	50
15	5	45	35	80	4(20)	50	50
11	9	41	39	80	4(20)	50	50

11. (2 pts) As we already know, *Agrobacterium tumefaciens* has a G+C content of 64%. Given no other information, which one of the following enzymes will probably cut the *A. tumefaciens* genome the most number of times?

DraI	5'-TTTAAA-3'
ApaI	5'-GGGCCC-3'
BamHI	5'-GGATCC-3'
EcoRI	5'-GAATTC-3'

fool-proof method:
probability of DraI cut = $(18/100)^6 = 0.00003$
probability of ApaI cut = $(32/100)^6 = 0.001$ WINNER!
probability of DraI cut = $(18/100)^2 \times (32/100)^4 = 0.0003$
probability of DraI cut = $(18/100)^4 \times (32/100)^2 = 0.0001$

quick estimate method:
Notice that all of the enzymes recognize a 6 bp sequence. Since G and C are the most likely nucleotides in the genome, the more G's and C's in the restriction site the higher the probability of finding the restriction site. (All of this assumes that nothing else influences the probability.)

12. (3 pts) You are working in the Emerging Diseases Unit, the phat section of the Center for Disease Control in Atlanta. Several people attending the 75th Annual Clown Convention in Baltimore have come down with a mysterious malady where their noses are bulbous and red without makeup!! No one has cultured the pathogen yet, but some intact

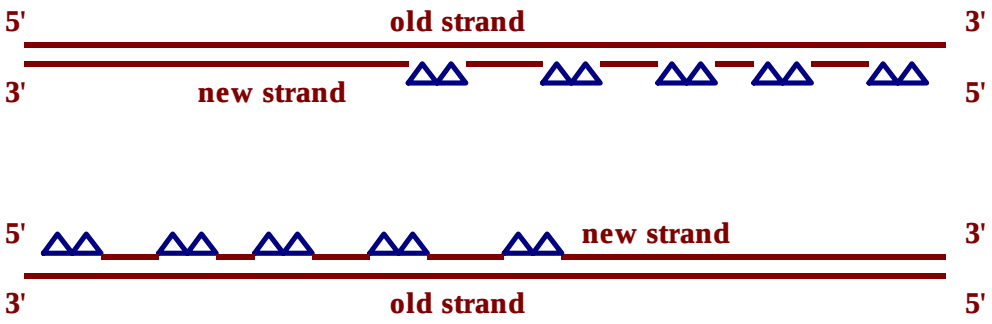
nucleic acid has been recovered from the noses of the victims that is not human in origin (don't ask me how we know that). You determine the nucleotide composition of the nucleic acid sample to be 20% A, 20% C, 30% G, and 30% T.

a. Is the nucleic acid DNA or RNA, single-stranded or double stranded?
T is present, so it is DNA. %A does not equal %T and %G does not equal %C, so it must be single-stranded.

b. Is the pathogen most likely a bacterium or a virus?
Cellular life forms and some viruses have double-stranded DNA genomes. However, other viruses have single-stranded DNA or RNA genomes. Our putative pathogen must be a virus.

13. (6 pts) A linear piece of DNA has just finished replicating from an origin of replication approximately in the center of the molecule. Draw a simple picture of the finished replication products in a mutant *E. coli* strain that lacks the 5'-to-3' exonuclease function of DNA Polymerase I. Provide a legend to your drawing and be sure to indicate the directionality of each strand. (If you have to make any assumptions to answer this question, please state them.)

Without the 5'-3' exonuclease activity of DNA Polymerase I, RNA primers cannot be degraded. Therefore, the new DNA strand are either lots of Okazaki fragments (if ligase cannot seal between DNA and RNA) or one continuous strand that is a hybrid of DNA and RNA. In the diagram below, DNA is red straight lines and RNA is blue wavy lines.



14. (15) ***E****. coli*, like many other bacteria, are able to grow in the complete absence of any amino acids in the growth medium -- they can make all the amino acids themselves. Two strains of *E. coli*, JC158 (Lac^+) and AB3517 (Lac^-), were mutagenized with ultraviolet radiation and colonies are screened for those that are unable to grow on minimal medium unless the medium is supplemented with the amino acid tryptophan. Three mutants were isolated in this way: Trp1, Trp2, and Trp3. The mutants have the following characteristics:

- When mutants Trp1 and Trp2 are spread on the appropriate indicator plates, they give rise to red colonies (owing to their Lac^+ phenotypes),

- When Trp3 is spread on the appropriate indicator plate, it gives cream-colored colonies.
- Chemical analysis of each mutant strain indicates that each excretes a characteristic chemical, as shown in Table 1.

Table 1: Characteristics of Trp⁻ mutants

Mutant	Parent	Lac phenotype	Chemical overproduced
Trp1	JC158	Lac^+	indole
Trp2	JC158	Lac^+	chorismate
Trp3	AB3517	Lac^-	anthranilate

Cultures of the three mutants are mixed together and plated on minimal medium plates (lacking tryptophan) that are supplemented with Lac indicator. The results are shown in Table 2.

Table 2: Results of mixing experiment

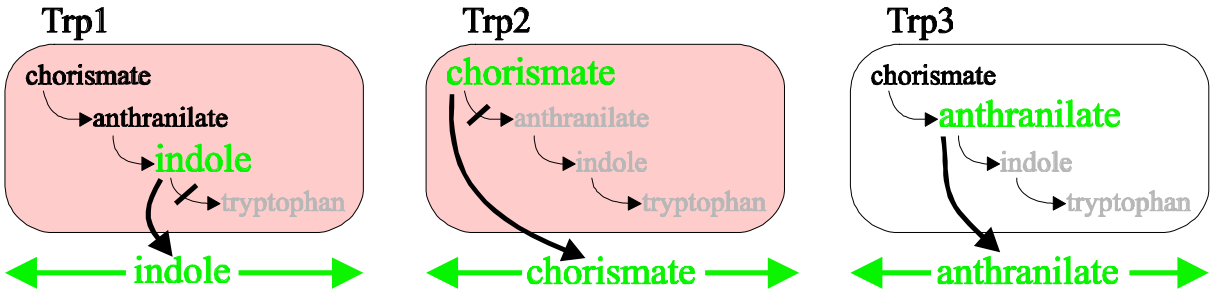
	Strain 1	Mixed with	Result
1	10^9 cells Trp1	---	No colonies
2	10^9 cells Trp2	---	No colonies
3	10^9 cells Trp3	---	No colonies
4	10^9 cells Trp1	100 cells Trp2	100 red colonies
5	10^9 cells Trp2	100 cells Trp1	No colonies
6	10^9 cells Trp1	100 cells Trp3	100 cream colonies
7	10^9 cells Trp3	100 cells Trp1	No colonies
8	10^9 cells Trp2	100 cells Trp3	No colonies
9	10^9 cells Trp3	100 cells Trp2	100 red colonies

14a. Which of the following are important factors in explaining the results:

A. DNA containing a *lac* gene is transferred between cells.
B. DNA containing a *trp* gene is transferred between cells.
C. An enzyme involved in lactose breakdown is transferred between cells.
D. An enzyme involved in tryptophan metabolism is transferred between cells.
E. Indole, chorismate, or anthranilate is transferred between cells.

DNA and protein don't pass between cells. They're way too large to get in and out past a hydrophobic membrane. Small molecules are often passed back and forth -- recall the case of embryonic flies and vermilion and cinnabar pigments.

14b. Using pictures and diagrams and as few words as possible, explain the results.



Each mutant is defective in the pathway shown above and releases the metabolite upstream from the block. The metabolite given off by Strain 1 of a mixture may be taken up Strain 2, and if the metabolite is downstream from the block of Strain 2, it may enter into the pathway and be transformed into tryptophan, allowing Strain 2 to grow and form colonies.

The situation is analogous to the transplantation of imaginal disks into fly abdomens. The pathway may be ordered by noting which metabolite in excess is able to feed into the pathway of the other strain. Strain 1 can never form colonies because there are too few cells of Strain 2 to support it. The color of the colonies always corresponds to the phenotype expected of Strain 2, confirm that it is always Strain 2 (if either) that grows.

15. (6) The DNA below contains the start codon of the gene encoding a toxin from wild-type *Clostridium botulinum* (the nontemplate[1] strand of DNA is shown). The bacterium is mutagenized with proflavin, and the result is a single-base insertion at the position shown below:

GCTGAGCTATGA↓GATCCTTAGTGACTAACCGTACG

15a. Do you believe that the mutant toxin will be:

A. Completely functional
B. Partially functional
C. Completely nonfunctional

The insertion of a base owing to the action of proflavin is downstream from the ATG start codon (starting at the ninth base of the sequence shown). Every codon beyond that point will be shifted by one base, hence the entire protein, except for the first amino acid, will be changed. You know that the ATG is the start codon for two reasons. First, the problem gives you that the strand shown is the NONtemplate strand. The footnote says that mRNA is complementary to the TEMPLATE strand, so the NONtemplate strand should have the same bases as the mRNA,… and there's only one ATG on that strand. Second, even if you choose to look on the other strand, there's no ATG there.

15b. What will be the last amino acid in the mutant protein?

Asp

The mutant protein is the one with the extra base put in by the action of proflavin. Start counting from the ATG: ATG - AxG - ATC - CTT - AGT - GAC - TAA… there's the stop codon! The codon before the TAA is GAC, which encodes Asp.

[1] The template strand is the strand used by RNA polymerase to make RNA and so is complementary to it.