Solutions to Genetics Problems

This chapter is much more than a solution set for the genetics problems. Here you will find details concerning the assumptions made, the approaches taken, the predictions that are reasonable, and strategies that you can use to solve any genetics problem. The value of this chapter depends on you. In every case, before you look here, you should struggle with the problem, design your own approach, and make your own mistakes. Only then should you look at the solutions given here.

(1) PROBLEMS INVOLVING ONLY ONE GENE

(1.1) One gene; two alleles; simple dominance

(1.1.1)
a) GG × GG. A plant homozygous for the G allele is crossed to another plant homozygous for the G allele. Each parent has only the G allele to give to its offspring, so the Punnett square used to predict the offspring would look like this:

	G	G
G	GG green	GG green
G	GG green	GG green

This predicts all offspring (100%) will have the genotype of GG and green flower color is the dominant trait, so all have green flowers.

Because both parents produce only one type of gamete, you can simplify the Punnett square to:

	G
G	GG green

b) gg × gg. A plant homozygous for the g allele is crossed to another plant homozygous for the g allele. Each parent has only the g allele to give to its offspring, so the Punnett square used to predict the offspring would look like this:

	g	g
g	gg blue	gg blue
g	gg blue	gg blue

or

	g
g	gg blue

This predicts all offspring (100%) will have the genotype of gg and blue flower color.

c) Gg × gg. A heterozygous plant is crossed to a plant homozygous for the g allele. The heterozygote will pass either the G or g allele to its offspring. The other parent has only the g allele to give. The Punnett square used to predict the offspring would look like this:

	G	g
g	Gg green	gg blue
g	Gg green	gg blue

or

	G	g
g	Gg green	gg blue

Because green flower color is dominant to blue flower color, the Gg offspring will have green flowers but the homozygous gg offspring will have blue flowers. Thus, you would expect 50% of the offspring to be Gg (green flowers) and 50% to be gg (blue flowers).

d) Gg × Gg. Two heterozygous plants are crossed. The Punnett square used to predict the offspring would look like this:

	G	g
G	GG green	Gg green
g	Gg green	gg blue

This predicts 25% GG offspring, 50% Gg offspring, and 25% gg offspring, a 1:2:1 genotypic ratio. However, because the Gg heterozygotes are indistinguishable from the GG homozygotes, what you see is 75% of offspring have green flowers and 25% of offspring have blue flowers, for a phenotypic ratio of 3:1.

e) Green × Green. There are two possible genotypes for a green individual: GG or Gg. This means that there are three possibilities for Green × Green:
1) GG × GG. This gives 100% GG with green flowers as in part (a) above.
2) Gg × Gg. This gives 75% green flowers (25% GG and 50% Gg) and 25% gg with blue flowers as in part (d) above.
3) GG × Gg. The Punnett square used to predict the offspring would look like this:

	G	g
G	GG green	Gg green
G	GG green	Gg green

This gives 50% GG and 50% Gg, but all offspring have green flowers.

Extra Challenge: Look at 1) and 3) above. If a green-flowered plant × a green-flowered plant gives all green-flowered offspring, how would you determine whether the parents were GG × GG or GG × Gg?

f) Blue × Blue
The only possible genotype for a blue individual is gg. Therefore, the cross must be gg × gg, which gives 100% gg (blue flowers) as in part (b) above.

(1.1.2)
a) In cross 1, a red-eyed mouse × a white-eyed mouse gives all red-eyed mice.
A possible model is that eye color is controlled by one gene with two alleles. Since the white-eyed phenotype is not seen in the offspring of this cross, it is likely that red eyes is the dominant phenotype. If so, cross 1 could be either:

A) red-eyed mouse (RR) × white-eyed mouse (rr) → all F_1 are Rr
B) red-eyed mouse (Rr) × white-eyed mouse (rr) → F_1 are either Rr or rr

Option (A) would give all red-eyed offspring, which is consistent with the observed results. Option (B) should give some white-eyed offspring with the genotype rr, so this does not fit the observed results.

In cross 2, a red-eyed offspring (F_1) × a red-eyed offspring (F_1) gives some red-eyed and some white-eyed offspring. We predict this cross to be: red-eyed mouse (Rr) × red-eyed mouse (Rr).

The expected results from this cross are:

	R	r
R	RR	Rr
r	Rr	rr

25% RR, 50% Rr, and 25% rr offspring, a 1:2:1 genotypic ratio. But in our model, the Rr heterozygotes are indistinguishable from the RR homozygotes, so what you see is that 75% of the offspring have red eyes and 25% of the offspring have white eyes, for a phenotypic ratio of 3:1.
The reported data of 36 red-eyed to 13 white-eyed fit this prediction well.

You could also have tried the alternative model, where white eyes are dominant. Some appropriate symbols are:

Allele	Contribution to phenotype
W	white eyes (dominant)
w	red eyes (recessive)

If so, cross 1 could be either:
A) red-eyed mouse (ww) × white-eyed mouse (Ww) where you would expect a 1:1 ratio of red eyes to white eyes. This is inconsistent with the data.

B) red-eyed mouse (ww) × white-eyed mouse (WW) where you would expect all Ww (white-eyed) offspring. This is also not observed, so this model is inconsistent with the data.

There is no need to check cross 2 since one piece of inconsistent data rules out this model. So, even if you did not conclude that red eyes were likely to be dominant, you can still propose a consistent model by ruling out models that do not fit the data.

Your complete model would be that eye color is controlled by one gene with two alleles where red eyes are dominant to white eyes.
Cross 1: RR × rr → all Rr
Cross 2: Rr × Rr → 36 (RR + Rr) and 13 (rr)

b) In cross 1, a long-eared mouse × a short-eared mouse gives some long-eared and some short-eared mice. A possible model is that ear length is controlled by one gene with two alleles. From these data, we cannot determine which allele is associated with the dominant phenotype, so we must look at the data from cross 2 before proposing a model.

In cross 2, a long-eared F_1 mouse × a long-eared F_1 mouse gives some long-eared and some short-eared mice. In this cross, long-eared parents produce mice with short ears, i.e., the short-eared phenotype was masked in the parents. Therefore, a likely model is that long ears is the dominant phenotype, and some appropriate symbols are:

Allele	Contribution to phenotype
L	long ears (dominant)
l	short ears (recessive)

Therefore, in cross 1, the long-eared parent could be LL or Ll, but the short-eared parent must be ll. So the cross could be either:

a) long-eared mouse (LL) × short-eared mouse (ll) where you would expect all (Ll) long-eared offspring. This is inconsistent with the data.

b) long-eared mouse (Ll) × short-eared mouse (ll) where you would expect a 1:1 ratio of (Ll) long ears to (ll) short ears. This is consistent with the data considering there are only 22 offspring to examine. 12:10 is approximately 1:1.

In cross 2, a long-eared F_1 mouse (Ll) × a long-eared F_1 mouse (Ll) should give offspring that have a ratio of three long-eared mice to one short-eared mouse. The data support this model.

You could also have tried an alternative model, where the short-eared phenotype is dominant. Some appropriate symbols are:

Allele	Contribution to phenotype
S	short ears (dominant)
s	long ears (recessive)

In this scenario, for cross 1 the short-eared parent could be SS or Ss, but the long-eared parent must be ss. So the cross could be either:

a) long-eared mouse (ss) × short-eared mouse (SS) where you would expect all short-eared offspring. This is inconsistent with the data.

b) long-eared mouse (ss) × short-eared mouse (Ss) where you would expect a 1:1 ratio of long ears to short ears. This is consistent with the data considering there are only 22 offspring to examine. 12:10 is approximately 1:1.

However, in cross 2, a long-eared F_1 mouse (ss) × a long-eared F_1 mouse (ss) should give only long-eared offspring, which is not seen. This model is inconsistent with the data.

So, even if you did not conclude that long ears were likely to be dominant, you can still propose a consistent model by ruling out models that do not fit the data.

(1.1.3)
a) The achondroplasia phenotype is dominant. If a novel phenotype that is not seen in the parents appears in their offspring, it suggests that the novel phenotype is recessive. By this reasoning, normal size is recessive to dwarf size, which is dominant. You can try the alternative model (dwarfism is recessive) and show that it is not consistent with these family data.

b) For the model that normal size is recessive to dwarf size, some appropriate symbols are:

Allele	Contribution to phenotype
D	dwarf (dominant)
d	normal height (recessive)

To have a dd (normal size) child, both parents must have at least one d allele. To be dwarves, they must both have at least one D allele. Thus, both parents must be Dd.

(1.1.4) Give all models that are consistent with the data.
a) red fly × red fly gives one blue fly progeny.
The model is: one color gene with two alleles where red color is dominant to blue color.
Appropriate symbols would be:

Allele	Contribution to phenotype
R	red (dominant)
r	blue (recessive)

The cross is red (Rr) × red (Rr) ⇒ blue (rr).

b) brown cow × white cow gives one brown cow progeny.
There are two possible models here:

1) There are two alleles of the color gene, and brown color is dominant to white color. Appropriate symbols would be:

Allele	Contribution to phenotype
B	brown (dominant)
b	white (recessive)

The cross would be brown (BB or Bb) × white (bb) ⇒ brown (Bb).

2) There are two alleles of the color gene, and white color is dominant to brown color. Appropriate symbols would be:

Allele	Contribution to phenotype
W	white (dominant)
w	brown (recessive)

The cross would be brown (ww) × white (Ww) ⇒ brown (ww).

The data are consistent with **both** of these models.

(V1)
a) There is no solution for this part.
b) i) TT × TT: the Punnett square predicts 100% TT; VGLII would give all three-wing and no five-wing.
ii) TT × Tt: the Punnett square predicts 50% TT and 50% Tt, both of which are three-winged; VGLII would give all three-wing and no five-wing.
iii) TT × tt: the Punnett square predicts 100% Tt; VGLII would give all three-wing and no five-wing.
iv) Tt × Tt: the Punnett square would predict 75% T_ (three-wing) and 25% tt (five-wing); in a VGLII problem, this would be observed as a mixture of three-wing and five-wing, with more three-wing than five-wing.
v) Tt × tt: the Punnett square would predict 50% Tt (three-wing) and 50% tt (five-wing); in a VGLII problem, this would be observed as a mixture of three-wing and five-wing, with roughly equal numbers of each.
vi) tt × tt: the Punnett square would predict 100% tt (five-wing); VGLII would give all five-wing and no three-wing.
c) There is no solution for this part.

(1.1.5) These mice have one gene with two alleles for the coat color trait. Brown is dominant to white. Some appropriate symbols would be:

Allele	Contribution to phenotype
B	brown (dominant)
b	white (recessive)

Parental cross:
brown mouse (Bb) × white mouse (bb) ⇒ 10 brown mice (Bb) and 13 white mice (bb)

White F_1 mice are bb. Therefore, in a cross between two white F_1 mice, all of their progeny will be white (bb) as well.

Brown F_1 mice are heterozygotes (Bb). The Punnett square used to predict the offspring would look like this:

	B	b
B	BB	Bb
b	Bb	bb

This predicts 25% BB offspring, 50% Bb offspring, and 25% bb offspring, a 1:2:1 genotypic ratio. However, because the Bb heterozygotes are indistinguishable from the BB homozygotes, what you expect is that 75% of the offspring should be brown and 25% of the offspring should be white. The results of 28 brown to 10 white are consistent with this model.

(1.1.6) There are many possible models. Below are two:

a) One gene with two alleles where blue is dominant to red.
i) BB = blue, Bb = blue, bb = red
ii)Tarzan (Bb) × Jane (bb) ⇒ Fred (bb) and Alice (bb)

b) One gene with two alleles where red is dominant to blue.

i) BB = red, Bb = red, bb = blue

ii) Tarzan (bb) × Jane (Bb) ⇒ Fred (Bb) and Alice (Bb)

c) i) Cross Fred × Alice.

ii) If model 1 is correct, then Fred (bb) × Alice (bb) would give all bb offspring which would be red.

iii) If model 2 is correct, then Fred (Bb) × Alice (Bb) would give 75% red offspring and 25% blue offspring.

(1.1.7)

a) Rr × Rr gives 1/4 chance of an rr child, and rr children have a 1/2 chance of being left-handed. Therefore, the chance is 1/4 x 1/2 or 1/8.

b) Yes. A left-handed (rr) mother × a right-handed (Rr) father have a 50% chance of having an rr child and that rr child could be left-handed.

c) Yes. Two left-handed parents, rr × rr, could have an rr child. That rr child could be right-handed.

d) This model allows for many possible individual families. However, on average, left-handed parents should be more likely to have left-handed children than right-handed parents are. When looking at many families, if right-handed parents were just as likely as left-handed parents to have left-handed children, the model is unlikely to be correct.

(V2) For this analysis, we will consider each cage in isolation. In a real analysis, you would take into account the related crosses you had done. You can check each of our assertions with a Punnett square, if you like.

a) Cage 2 is consistent only with Model A. If Model B were correct, it would not be possible to have yellow parents produce black offspring. Cage 5 is similarly conclusive.

b) Considered on their own, all the other cages give inconclusive results. That is:

- Cage 3: black × black giving all black offspring is consistent with either Model A or B.
- Cage 4: yellow × black giving a mixture of black and yellow offspring is consistent with either Model A or B.

Although these are not conclusive on their own, in context, they support our model.

c) Another result that would be conclusive is if you crossed black × yellow and got only yellow offspring. This is not consistent with Model B.

d) There is no answer for this part.

e) There is no answer for this part.

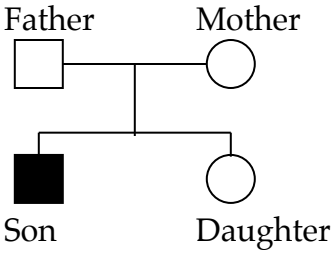
(1.2) Pedigrees involving one gene, I

(1.2.1)

a) i) You could use almost any letter for the cystic fibrosis gene; here is one example:

Allele	Contribution to phenotype
F	normal (dominant)
f	cystic fibrosis (recessive)

ii) Since pedigree symbols show only phenotype, the symbols for a carrier and a homozygous normal individual are identical. Therefore, the pedigree would look like this:



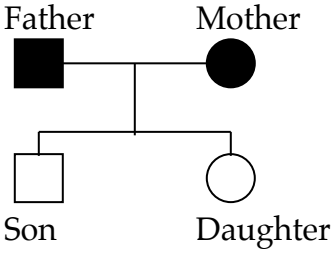
iii) From the information in the problem, the parents are carriers: Ff. The son with cystic fibrosis is ff. Using a Punnett square for these parents, the unaffected daughter could be either FF or Ff; this can also be written as F_. Without more information, it is not possible to know her genotype for certain.

iv) Using the Punnett square for these parents, 1/4 of the children, on average, would be expected to have cystic fibrosis. Therefore, the chance that the next child would have cystic fibrosis is 1/4 or 25%.

b) i) You could use almost any letter for the Marfan syndrome gene. Here is one example:

Allele	Contribution to phenotype
M	Marfan syndrome (dominant)
m	normal (recessive)

ii) The pedigree would look like this:



iii) In the case of an autosomal dominant trait, normal individuals can only be mm (if they had even one M allele, they would have Marfan syndrome). So the normal son must be mm.

iv) Without knowing anything about their children, the parents with Marfan syndrome could be either MM or Mm. You can figure out which genotype they have by considering their offspring. There are two ways to do this:

1) Try all three possibilities and see which can produce an mm (normal) child. If you try MM × MM, MM × Mm, or Mm × MM, none of these can produce an mm offspring. However, Mm × Mm can produce mm offspring. Therefore, both parents must be Mm. You might be surprised to see two out of two (100%) normal children when the Punnett square predicts only 1/4 of the children of these parents to be normal. This is not surprising since the number of offspring is small and statistical fluctuations are to be expected.

2) Work backward from the offspring. Consider the unaffected son; he must be mm. Since he got one of his alleles from his mother and one from his father, both mom and dad must have at least one m allele. Since mom and dad have Marfan syndrome, they each must have at least one M allele. Combining these, the parents must both be Mm.

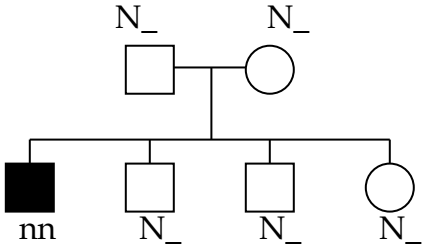
Either way, the parents must be Mm. Using a Punnett square, 3/4 of their offspring, on average, will have Marfan syndrome. So the risk that the next child in this family will have Marfan syndrome is 3/4 or 75%.

(1.2.2)
a) You can try both possibilities and see which works.

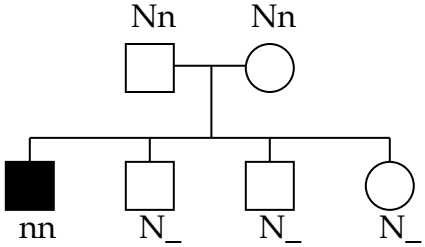
1) If having the disease is a recessive phenotype, then some appropriate symbols are:

Allele	Contribution to phenotype
N	normal (dominant)
n	diseased (recessive)

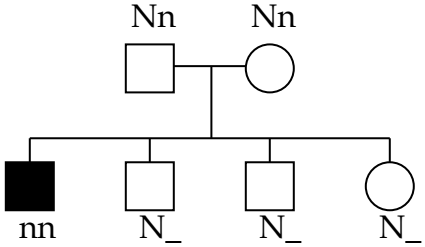
Start by writing the genotypes you are SURE of – the ones you can tell by phenotype alone. If the disease is inherited as an autosomal recessive trait, then you know that any diseased individuals must be nn. You also know that normal individuals must have at least one N in order to be normal. They could be either NN or Nn; you cannot be sure which without more information.



Then, work from what you know and see if the inheritance is possible. One place to start is the diseased son. He had to get an n from both parents, so they must be Nn:



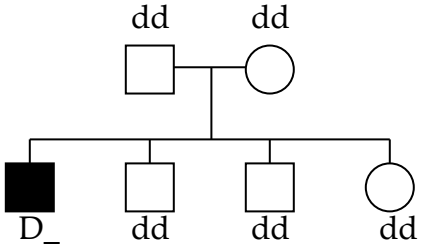
Finally, evaluate the other children. Is it possible for two Nn parents to have unaffected children with at least one N allele? The answer is yes, and without additional information, the genotype of the three unaffected children remains ambiguous. They could be either NN or Nn, so this model is consistent with the data. The genotypes are as follows:



2) On the other hand, if having the disease is a dominant phenotype, then some appropriate symbols are:

Allele	Contribution to phenotype
D	disease (dominant)
d	normal (recessive)

Start by writing the genotypes you are SURE of – the ones you can tell by phenotype alone. If the disease is inherited as an autosomal dominant trait, then you know that any normal individuals must be dd. You also know that diseased individuals must have at least one D in order to be diseased. They could be either DD or Dd; you cannot be sure which without more information.



Now, evaluate the pedigree and see whether the inheritance is possible. One place to start is the diseased son. He had to get a D from one of his parents, so at least one of them must have a D allele. However, if one of his parents had a D, he or she would be diseased. This is inconsistent with the pedigree so this disease cannot be inherited in an autosomal dominant manner.

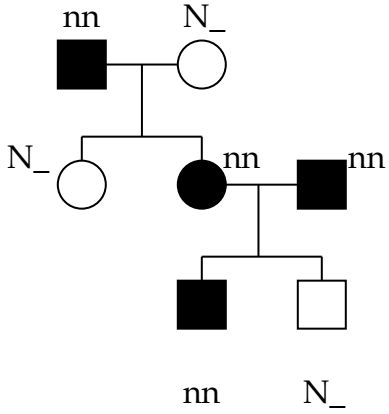
For part (a) above, we've shown that the pedigree is consistent with the disease being inherited as an autosomal recessive trait and inconsistent with the disease being inherited as an autosomal dominant trait. Therefore, the disease must be inherited in an autosomal recessive manner.

b) You can try both possibilities and see which works.

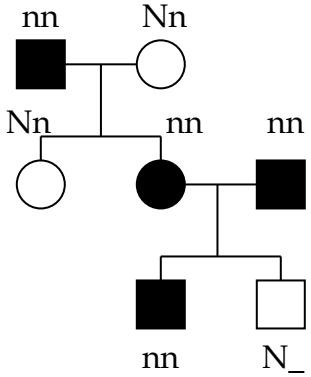
1) If having the disease is a recessive phenotype, then some appropriate symbols are:

Allele	Contribution to phenotype
N	normal (dominant)
n	diseased (recessive)

Start by writing the genotypes you are SURE of – the ones you can tell by phenotype alone. If the disease is inherited as an autosomal recessive trait, you know that any diseased individuals must be nn and normal individuals are N_.



Start from any of the diseased individuals. Try from the top down. To have a diseased daughter, the grandmother (top row; right) must have an n allele. It is possible for the grandparents (nn × Nn) to have both normal (Nn) and diseased (nn) offspring, so the first generation is consistent with the disease being a recessive trait.

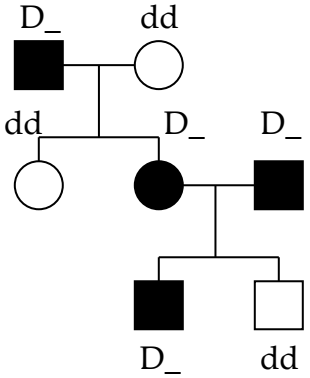


Now examine the second generation. The parents are nn × nn. The only offspring they can have are nn, so it is not possible for them to have normal children. The normal child at the lower right is inconsistent with this model. So this disease cannot be inherited in an autosomal recessive manner.

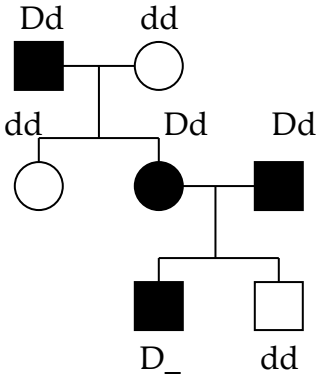
2) The alternative is that the disease is a dominant phenotype, and some appropriate symbols are:

Allele	Contribution to phenotype
D	diseased (dominant)
d	normal (recessive)

Start by writing the genotypes you are SURE of – the ones you can tell by phenotype alone. If the disease is inherited as an autosomal dominant trait, you know that any normal individuals must be dd and diseased individuals are D_.



You could start from any of the diseased individuals. To have a diseased daughter, one or both of the grandparents (top row) must have a D allele and show the disease. In addition, to have a normal daughter, both of the grandparents (top row) must have at least one d allele. Thus, the grandparents could be Dd × dd and be consistent with the disease as an autosomal dominant trait.



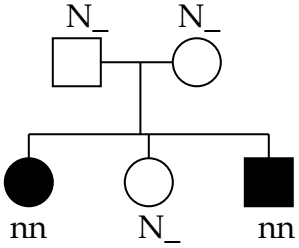
Now try the second generation. The parents are Dd × Dd. They could have DD, Dd, or dd offspring, so this pedigree is consistent with the disease being inherited as a dominant trait.

c) You can try both possibilities and see which works.

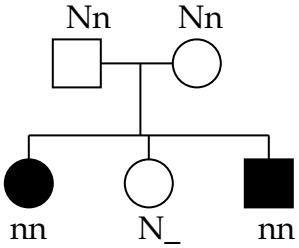
1) If having the disease is a recessive phenotype, then some appropriate symbols are:

Allele	Contribution to phenotype
N	normal (dominant)
n	diseased (recessive)

Start by writing the genotypes you are SURE of – the ones you can tell by phenotype alone. If the disease is inherited as an autosomal recessive trait, then you know that any diseased individuals must be nn. You also know that normal individuals must have at least one N in order to be normal. They could be either NN or Nn; you cannot be sure which without more information.



You know the parents must each have an n in order to produce nn children, so the parents must be Nn:



Then ask if Nn × Nn can produce normal offspring. They can have normal offspring that are Nn or NN. So this pedigree is consistent with the disease being inherited as a recessive trait.

2) If the disease is a dominant phenotype, then each generation must show the disease. In this pedigree, normal parents have diseased children, so the disease cannot be a dominant trait.

(1.2.3) a) Are the following statements true for autosomal recessive and/or autosomal dominant diseases:

i) *Diseased parents can have diseased offspring*. Yes, this is true for both autosomal dominant and autosomal recessive. Using the allele symbols defined in problem (1.2.2):

- Autosomal recessive: Two diseased parents (nn × nn) can have diseased offspring.
- Autosomal dominant: Two diseased parents (Dd × Dd) can have diseased offspring.

ii) *Normal parents can have normal offspring*. Yes, this is true for both autosomal dominant and autosomal recessive. Using the allele symbols defined in problem (1.2.2):

- Autosomal recessive: Two normal parents (Nn × Nn) can have normal offspring.
- Autosomal dominant: Two normal parents (dd × dd) can have normal offspring.

iii) *Even if both parents are normal, they can have diseased offspring*. This is only true for autosomal recessive.

- Autosomal recessive: Two normal parents (Nn × Nn) can have diseased offspring.
- Autosomal dominant: Two normal parents (dd × dd) cannot have normal offspring.

iv) *Even if both parents are diseased, they can have normal offspring*. This is true only for autosomal dominant.

- Autosomal recessive: Two diseased parents (nn × nn) cannot have normal (N_) offspring.
- Autosomal dominant: Two diseased parents (Dd × Dd) can have normal offspring.

b) The last two statements are diagnostic for particular modes of inheritance.

If you have two normal parents that have one or more diseased children, the disease cannot be inherited in an autosomal dominant manner. This can be used in problem (1.2.2), a and c, to rule out the autosomal dominant mode of inheritance. Note that you still have to check to be sure that autosomal recessive works.

If you have two diseased parents that have one or more normal children, the disease cannot be inherited in an autosomal recessive manner. This can be used in problem (1.2.2) b to rule out autosomal recessive. Note that you still have to check to be sure that autosomal dominant works.

(1.2.4) Fred would be at greater risk. Fred has an affected sister (dd) but his parents are normal so Fred's parents must be Dd × Dd. Therefore, the risk that Fred will be diseased is 1/4.

John's mother is dd, and his Dad is normal (DD or Dd). If Dad is DD, John cannot be affected; if Dad is Dd, John has a 1/2 chance of being affected. But because this is a rare disease, the chance that John's Dad is a carrier is very low. The actual risk can be calculated as 1/2 × (the chance that Dad is a carrier).

(1.2.5)
a) There is no right or wrong answer here, but it looks like Marfan syndrome runs in Anne's father's family. It appears that Charlie, John, and Peter have Marfan syndrome. This is consistent with autosomal dominant inheritance where affected children must have affected parents. It is also likely that Anne and probably David have Marfan syndrome.

b) Since Marfan syndrome is inherited in an autosomal dominant manner, children of an affected parent have a 50% chance of having Marfan syndrome.

(1.3) One gene; more complex models, I

(1.3.1) In *incomplete dominance*, the heterozygote has an *intermediate* phenotype, a phenotype different from either of the homozygotes. Some appropriate symbols would be:

Genotype	Phenotype
TT	tall
T′T′	short
TT′	medium, this is intermediate between tall and short

a) The parents would be TT × T′T′, giving all TT′ (medium height) offspring.

b) The parents can only be TT′ × TT′ giving:
25% TT – tall
25% T′T′ – short
50% TT′ – medium

(1.3.2) In *codominance*, the heterozygote has a *mixture of both* homozygote phenotypes. For example:

Genotype	Phenotype
LL	long hair
L′L′	short hair
LL′	a mixture of both long and short hair

a) The parents would be LL × L′L′, giving all LL′. The heterozygote offspring would have a mix of long and short hair.

b) The parents can only be LL′ × LL′, giving:
25% LL – long hair only
25% L′L′ – short hair only
50% LL′ – mixed long and short hair

(1.3.3)

Genotype	Phenotype
$C^B C^B$	blue
$C^R C^R$	red
c c	green
$C^B C^R$	blue
C^B c	blue
C^R c	red

(1.3.4) In cross 1, a blue-flowered plant × a white-flowered plant gives offspring that all have pale blue flowers. A plausible model is that color is controlled by one gene with two alleles and that color is incompletely dominant such that:

genotype	phenotype
BB	blue
BB′	pale-blue
B′B′	white

If so, cross 1 is: BB × B′B′ ⇒ all BB′, pale blue.

This would predict that cross 2 is: BB′ × BB′ ⇒ 1:2:1 blue flowers (BB) : pale blue flowers (BB′) : white flowers (B′B′).

For both cross 1 and cross 2, the predictions agree with the data.

(1.3.5) If you look at only cross 1, you see two phenotypes, green-eyed and white-eyed, so you could try a two-allele model. Cross 1 also tells us that we have heterozygote parents with green eyes that have some offspring with white eyes. This would indicate that green eyes are dominant to white eyes. You could use the symbols:

Genotype	Phenotype
GG	green eyes
Gg	green eyes
gg	white eyes

Cross 1 would have been Gg × Gg. You would then predict 25% GG, 50% Gg, and 25% gg or a ratio of 3 green-eyed : 1 white-eyed insects in the offspring. This is consistent with cross 1.

If you consider cross 2, however, you see three phenotypes. Both incomplete dominance and more than two alleles could explain three phenotypes, so where do you begin? If the three eye colors were due to incomplete dominance, you would expect to have seen red eyes in the first cross; the parents were not homozygotes because you had a mix of eye colors in the offspring. So you should consider the possibility that eye color is controlled by one gene with three alleles.

a) i) Thus, the three eye colors could be due to three alleles of the eye color gene where green eyes are dominant to white eyes. Some appropriate symbols would be:

E^R – allele associated with red eyes
E^G – allele associated with green eyes
E^W – allele associated with white eyes

ii) The parents in cross 1 would be: $E^GE^W \times E^GE^W$

iii)

	E^G	E^W
E^G	E^GE^G	E^GE^W
E^W	E^GE^W	E^WE^W

You would then predict 25% E^GE^G, 50% E^GE^W and 25% E^WE^W, or a ratio of 3 green-eyed : 1 white-eyed insects in the offspring. This is consistent with cross 1.

b) i) In cross 2, red-eyed × white-eyed → red-eyed and green-eyed offspring.

- The white-eyed phenotype is masked in the offspring so assume white eyes are recessive to both red and green eyes and thus the white-eyed parent would be: E^WE^W.
- Both red and green eyes are seen in the offspring; thus the red-eyed parent must have both the E^R and the E^G alleles, which means that red eyes are dominant to green eyes.

ii) Therefore, the parents in cross 2 are: red-eyed (E^RE^G) × white-eyed (E^WE^W).

(1.3.6) To begin, assign alleles to each of the parents. If both parents' genotypes are unambiguous, then predict the blood types possible in their offspring. If the parental genotypes are ambiguous, then predict blood types possible in the offspring for each combination.

i) type AB = I^AI^B , type O = ii: these parents could have I^Ai (type A) or I^Bi (type B) children.

ii) type A = I^AI^A or I^Ai, type O = ii.

If the type A parent is I^AI^A, then the couple could have only I^Ai (type A) children.

If the type A parent is I^Ai, then the couple could have I^Ai (type A) or ii (type O) children.

iii) type A = I^AI^A or I^Ai, type AB = I^AI^B.

If the type A parent is I^AI^A, then the couple could have I^AI^A (type A) or I^AI^B (type AB) children.

If the type A parent is I^Ai, then the couple could have I^AI^A (type A), I^AI^B (type AB), I^Ai (type A), or I^Bi (type B) children.

iv) type O = ii: these parents could have only ii (type O) children.

- Couple (iv) could have had only the baby with blood type O.
- The baby with type AB blood could have come only from couple (iii).
- Since couple (iii) had the AB baby, then the child with type B blood belongs to couple (i).
- This leaves the child with type A blood belonging to couple (ii).

(1.3.7) Remember a child receives one and only one allele from each parent.

- If George and Sallie are indeed Fred's parents, then Fred (type B blood) must have received the I^B allele from his father. Sallie with type A blood does not have the I^B allele.

- If George and Sallie are Fred's parents, then Sallie must have the genotype I^Ai, and Fred would have gotten the i allele from Sallie.

a) With only the blood type information, George and Sallie could be Fred's parents.

b) The information that George's father has type A blood and his mother has type B blood restricts George's genotype to I^Bi. This is still consistent with George and Sallie being Fred's parents.

c) The information that George has a sister with type O blood defines George's father as I^Ai and his mother as I^Bi, but this does not change the possibility that George and Sallie are Fred's parents.

d) The information that Sallie's father and mother are both I^AI^B means that Sallie (type A) has the genotype I^AI^A. This would prevent her from giving Fred the i allele. So if all the family information is true, then George and Sallie cannot be Fred's parents.

(1.3.8) The mother and the child are both type O and must have the genotype ii.

a) Bob with type A blood could be I^AI^A or I^Ai. If he is I^Ai, he could contribute the i allele, so he cannot be ruled out as the child's father.

b) Bob's mother (type A) could be I^AI^A or I^Ai and his father (type AB) must be I^AI^B. Bob could have type A blood (and the I^Ai genotype) if his mother contributed her i allele and his father contributed his I^A allele. Therefore, this information cannot exclude the man as the child's father.

c) If Bob's mother's parents are both type AB (I^AI^B), then Bob's mother must be I^AI^A, and she could not contribute an i allele to her son. Therefore, Bob must also be I^AI^A. This information would exclude him as the child's father.

(1.3.9) Begin by assigning the parental genotypes and the potential blood types of the children.

a) Couple #1 cannot be Rodger's parents.

> Tom must be I^AI^B and Ann is I^AI^A or I^Ai.
> If Ann is I^AI^A, then their children could be type A (I^AI^A) or type AB (I^AI^B).
> If Ann is I^Ai, then their children could be type A (I^AI^A or I^Ai), type AB (I^AI^B), or type B (I^Bi). But they cannot have a type O child. Therefore, they are not Rodger's parents.

Couple #2 cannot be Cathy's parents.

> Peter and Sally are both either I^BI^B or I^Bi. They can have type B children and, if they are both I^Bi, a type O child. They cannot have a type A child. Therefore, they are not Cathy's parents.

b) Since Ann's parents are both I^AI^B, she must be I^AI^A. Therefore, Ann and Tom (I^AI^B) can have only type A or AB children; they cannot have a type B child. In (a) you determined that they cannot have a type O child; therefore, they must be Cathy's parents.

c) Since Peter's parents are I^B_ and ii, he must be I^Bi. Since Sally's parents are both I^AI^B, she must be I^BI^B. Peter and Sally thus can have a type B child, but they cannot have a type O child. Therefore, Peter and Sally are Steve's parents.

(V3) There are no solutions for this part.

(1.3.10)

a) In cross 2, a purple plant is crossed to a blue plant and all the offspring are purple.

> i) Assume that color is controlled by one gene with two alleles, and that purple color is dominant to blue color. Some appropriate symbols would be:
>
> | Genotype | Phenotype |
> |---|---|
> | PP | purple |
> | Pp | purple |
> | pp | blue |
>
> ii) Cross 1 = purple (Pp) × blue (pp). Our model predicts 50% purple (Pp) and 50% blue (pp) offspring, which is what the data show.
> Cross 2 = purple (PP) × blue (pp). Our model predicts all purple (Pp) offspring, which is what the data show.

b) In these plants, you see three phenotypes. Both incomplete dominance and more than two alleles could explain three phenotypes, so where do you begin? If the three colors were due to incomplete dominance, then you could predict that the purple phenotype (which is intermediate between the blue and the red phenotypes) is associated with the genotype Pp. If this were true, then you would expect cross 3 and cross 4 to give identical results. Therefore, you should consider the possibility that color is controlled by one gene with three alleles.

Because crosses 3 and 4 give different offspring, you know that at least one parent in each cross is a heterozygote. However, all the parents in crosses 3 and 4 are purple, so purple is dominant to both red and blue.

Cross 5 indicates that blue is dominant to red.

Our model is that color is controlled by one gene with three alleles. Purple is dominant to both red and blue, and blue is dominant to red. Some appropriate symbols would be:

C^R – allele associated with red
C^B – allele associated with blue
C^P – allele associated with purple

c)
Cross 1 purple parent = C^PC^B
Cross 2 purple parent = C^PC^P
Cross 3 purple parent = C^PC^B
Cross 4 purple parent = C^PC^R
Cross 5 purple parent = C^BC^B

Therefore:
Cross 3 = purple (C^PC^B) × purple (C^PC^B) ⇒ 3 purple (C^PC^P or C^PC^B) : 1 blue (C^BC^B)
Cross 4 = purple (C^PC^R) × purple (C^PC^R) ⇒ 3 purple (C^PC^P or C^PC^R) : 1 red (C^RC^R)
Cross 5 = blue (C^BC^B) × red (C^RC^R) ⇒ all blue (C^BC^R)

(1.3.11)
You see four coat colors in tribbles. Both incomplete dominance and more than two alleles could explain these phenotypes. Crosses 1 and 2 do not indicate that the colors are due to incomplete dominance. In fact, given crosses 1 and 2, you would predict that color is controlled by one gene with three alleles and that green is dominant to both red and white. In the progeny from cross 3 (where your F_1 tribbles are likely heterozygotes), you see what could be an intermediate phenotype.

Therefore, you should consider the possibility that color is controlled by one gene with three alleles, but some colors show incomplete dominance.

a) Our model is that color is controlled by one gene with three alleles. Green is dominant to both red and white, but the red and white phenotypes show incomplete dominance with each other. Some appropriate symbols would be:

Genotype	Coat color
$C^G\_$	green
C^RC^R	red
C^WC^W	white
C^RC^W	pink

The green F_1 tribbles from cross 1 are heterozygous C^GC^R. The green F_1 tribbles from cross 2 are heterozygous C^GC^W.
So cross 3 = green (C^GC^R) × green (C^GC^W) ⇒25% C^GC^G (green) : 25% C^GC^R (green) : 25% C^GC^W (green) : 25% C^RC^W (pink), giving the ratio of 3 (green) : 1 (pink) seen.

b) The cross is pink (C^RC^W) × green (C^GC^G) ⇒ 50% C^RC^G and 50% C^WC^G. Because green is dominant to both white and red, all progeny will be green.

(1.4) One gene; sex linkage

(1.4.1)

a) We are following an X-linked gene where red eyes is dominant to white eyes.

i) The white-eyed female can only be X^rX^r; the red-eyed male can only be X^RY.

The Punnett square is:

	X^r	X^r
X^R	X^RX^r red-eyed	X^RX^r red-eyed
Y	X^rY white-eyed	X^rY white-eyed

This would give:

50% X^RX^r red-eyed female
50% X^rY white-eyed male

ii) The red-eyed female can be X^RX^r or X^RX^R; the white-eyed male can only be X^rY.

The first Punnett square is:

	X^R	X^r
X^r	X^RX^r red-eyed	X^rX^r white-eyed
Y	X^RY red-eyed	X^rY white-eyed

This would give:

25% X^RX^r red-eyed female
25% X^rX^r white-eyed female
25% X^RY red-eyed male
25% X^rY white-eyed male

The second Punnett square is:

	X^R	X^R
X^r	X^RX^r red-eyed	X^RX^r red-eyed
Y	X^RY red-eyed	X^RY red-eyed

This would give:

50% X^RX^r red-eyed female
50% X^RY red-eyed male

b) We are following a Z-linked gene where red eyes are dominant to white eyes.

i) The white-eyed female can only be Z^rW; the red-eyed male can be either Z^RZ^r or Z^RZ^R.

The first Punnett square is:

	Z^r	W
Z^R	Z^RZ^r red-eyed	Z^RW red-eyed
Z^r	Z^rZ^r white-eyed	Z^rW white-eyed

This would give:

25% Z^RW red-eyed female
25% Z^rW white-eyed female
25% Z^RZ^r red-eyed male
25% Z^rZ^r white-eyed male

The second Punnett square is:

	Z^r	W
Z^R	Z^RZ^r red-eyed	Z^RW red-eyed
Z^R	Z^RZ^r red-eyed	Z^RW red-eyed

This would give:

50% Z^RW red-eyed female
50% Z^RZ^r red-eyed male

ii) The red-eyed female can only be Z^RW; the white-eyed male can only be Z^rZ^r.

The Punnett square is:

	Z^R	W
Z^r	Z^RZ^r red-eyed	Z^rW white-eyed
Z^r	Z^RZ^r red-eyed	Z^rW white-eyed

This would give:

50% Z^rW white-eyed female
50% Z^RZ^r red-eyed male

(1.4.2)
a) A combination that is inconsistent with autosomal recessive.
Two diseased parents having a normal child. The parents must both be nn, so the child can only be nn (diseased).

b) Combinations that are inconsistent with sex-linked recessive.

- Two diseased parents having a normal child. The father must be X^nY and the mother must be X^nX^n, so the sons will all be X^nY (diseased) and the daughters will all be X^nX^n (diseased).
- A normal father having a diseased daughter. The daughter must be X^nX^n. Therefore, she must have gotten an X^n from each parent. Therefore, the father has to be X^nY (diseased).
- A diseased mother having a normal son. The mother must be X^nX^n. Since the son gets his X from his mother, he must be X^nY (diseased).

c) A combination that is inconsistent with autosomal dominant.
A diseased child from two normal parents. The child has to have at least one D allele. That allele had to come from one of the parents. That parent would therefore have to have at least one D, which would make him/her diseased.

(V4) There are no solutions for this part.

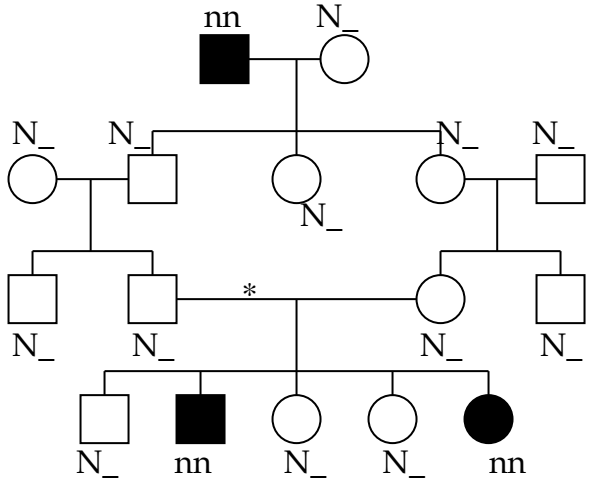
(1.5) Pedigrees involving one gene, II

(1.5.1)
a) There are two ways to solve parts (i) and (ii) of this problem. The first is the "brute force" method: try all three possible genetic models. This is time-consuming but guaranteed to get you the right answer.

1) First, try <u>autosomal recessive</u>:

<u>Genotype</u>	<u>Phenotype</u>
NN or Nn	normal
nn	affected

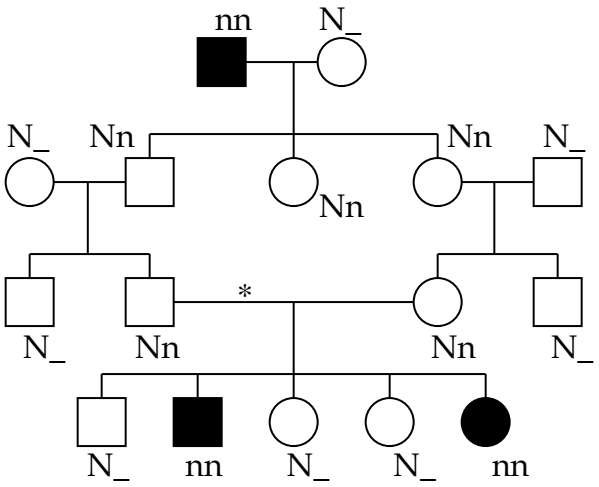
As before, assign the genotypes you know for sure. Individuals whose genotypes are ambiguous can be labeled as N_.



Begin with an affected individual (nn) on the bottom row; that individual has to get one (n) allele from each parent, so all ambiguous parents of affected individuals become (Nn). Thus, the parents of the affected son in the bottom row can be assigned as Nn. What about his siblings? It is possible for two Nn parents to have both affected and normal offspring, so without further information all his normal siblings remain (N_).

Now look at the affected male in the top row. All of his children have to get an n allele from him (since that is all he has to give). We don't know what the genotype of the top row female is, so she has to be marked as N_. (*Note that since the trait is rare, she is more likely to be NN.)

The pedigree incorporating this information is below.

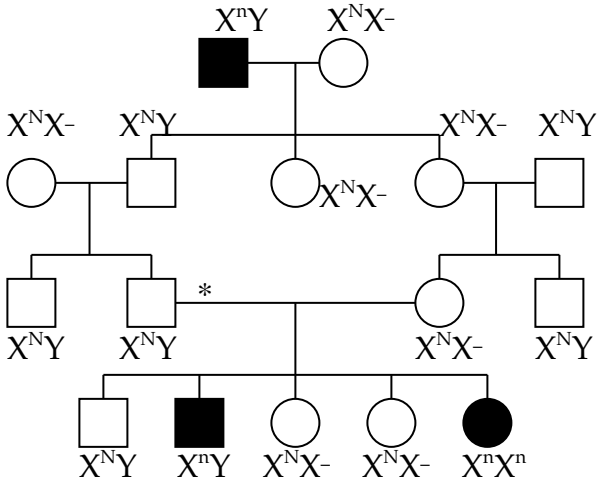


We cannot assign genotypes to the other individuals in the pedigree, so they must be marked N_. You can check to see that all the combinations of parents and offspring are possible. Therefore, this trait could be inherited in an autosomal recessive mode.

2) Now try sex-linked recessive:

Genotype	Phenotype
X^NX^N	normal female
X^NX^n	normal female
X^nX^n	affected female
X^NY	normal male
X^nY	affected male

This gives:

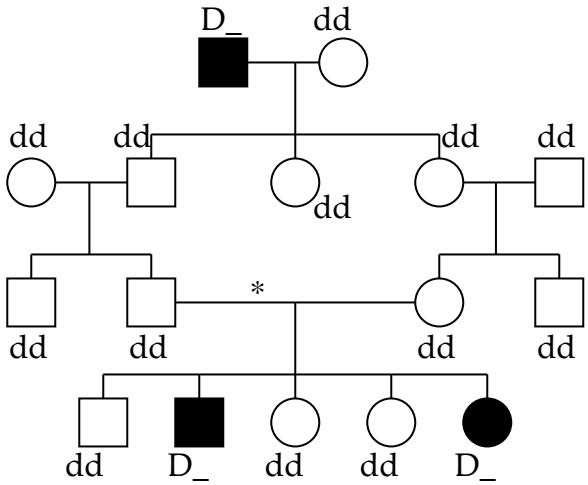


Begin with the affected son in the bottom row: he got his Y from his dad and his X^n from his mom. Therefore, his mom has to be X^NX^n. What about his affected sister? She had to get an X^n from her mom and an X^n from her dad. But if her dad had an X^n, he would have to be affected and he isn't. Therefore this pedigree is not consistent with a trait that is inherited in a sex-linked recessive manner.

3) Now try autosomal dominant:

Genotype	Phenotype
DD or Dd	affected
dd	normal

This gives:



Look at the affected son in the bottom row. One or more of his parents must have at least one D allele. But that would make them affected and they are normal, so this trait cannot be inherited in an autosomal dominant manner.

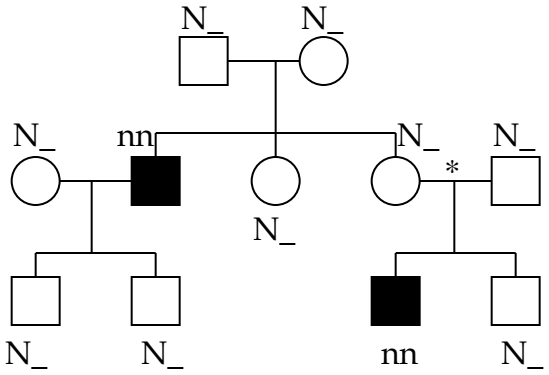
Here we tried all possible modes of inheritance. By process of elimination, this disease is consistent only with autosomal recessive inheritance.

iii) The couple marked with a * are Nn × Nn, so the risk that the next son will be diseased is 1/4 or 25%.

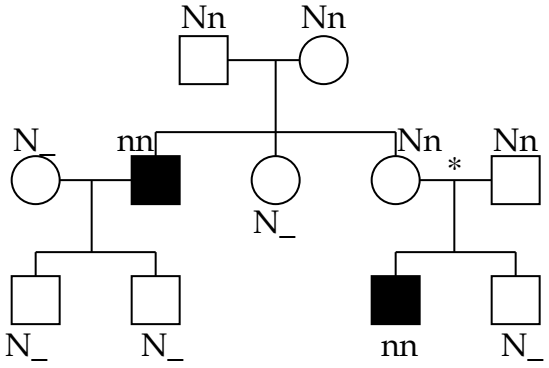
Another way to approach this pedigree is to use the rules that you created in an earlier problem (1.4.2) to rule out some of the models. Using those rules, the affected daughter in the bottom row eliminates two models. First, since her father is normal, the trait cannot be sex-linked recessive. Second, since both of her parents are normal, the trait cannot be autosomal dominant. You could then check to see if autosomal recessive works and it does.

b) Using the rules outlined in problem (1.4.2), we can immediately rule out autosomal dominant because both affected children have two normal parents. That leaves autosomal recessive and sex-linked recessive. No parts of the pedigree rule either of those two models out, so we must try each one to see how it works.

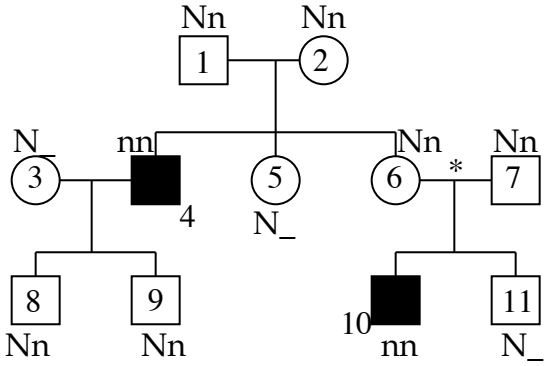
1) Try an autosomal recessive mode of inheritance. First fill in the genotypes we know for sure. All affected individuals would be nn and all unaffected individuals would have at least one N.



Begin with the affected male in the bottom row. He had to get an n from both parents. So they have to be Nn. Two Nn parents can have both normal and diseased children, so this part of the pedigree is consistent with an autosomal recessive mode of inheritance. Now look at the affected male in the middle row. As before, both of his parents have to be Nn. This information has been incorporated in the following pedigree.



Now look again at the affected male in the middle row. He must give an (n) to each of his children. Since none of his children are affected, they must be Nn. We do not have any other information that allows us to assign genotypes to the remaining individuals, so if the trait were inherited as an autosomal recessive trait, then the pedigree would be as follows.



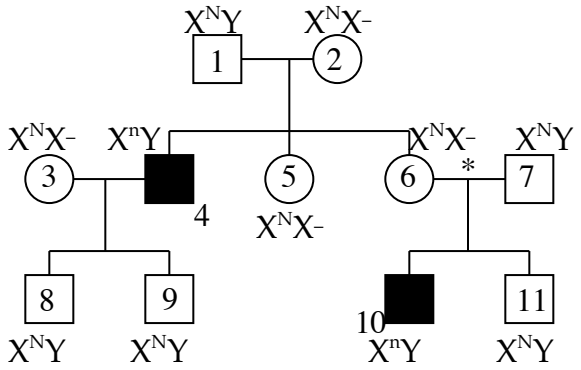
For the purposes of assessing how likely this mode of inheritance is we need to count the unrelated carriers; that is, unrelated people who have at least one allele associated with the trait. In this pedigree, there are three unrelated people with the trait allele: 1, 2, and 7. All the other people who have one or more n alleles got their n alleles from individuals 1, 2, or 7 or their descendants.

Going through each of the individuals:

1) Brought his n into the family. Unrelated carrier #1
2) Brought her n into the family. Unrelated carrier #2
3) Does not need to have an n to make the pedigree work.
4) Got his n's from 1 and 2.
5) If she has an n, she got it from 1 or 2.
6) Got her n from 1 or 2.
7) Brought his n into the family. Unrelated carrier #3
8) Got his n from 4.
9) Got his n from 4.
10) Got his n's from 6 and 7.
11) If he has an n, he got it from 6 or 7.

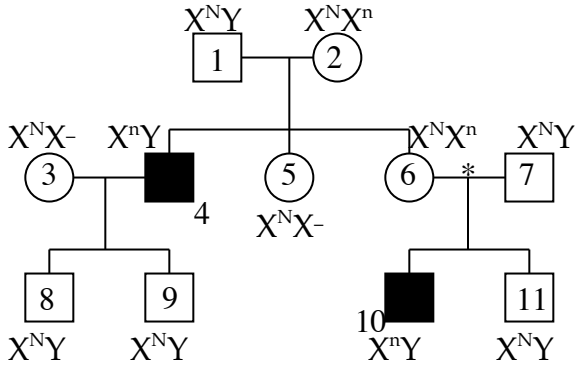
So for this trait to be inherited as an autosomal recessive trait, three unrelated carriers are required.

2) Now try a sex-linked recessive mode of inheritance; first fill in the genotypes you know for sure:



Begin with the affected male in the bottom row, individual 10. He got his Y from dad and his X^n from mom, so mom has to be X^NX^n. It is possible for $X^NX^n \times X^NY$ to have affected and normal sons, so this part of the pedigree is consistent with a sex-linked recessive mode of inheritance.

If you look at individual 4, his mom also has to be X^NX^n. Given this, the siblings of individual 4 are also possible so the pedigree remains consistent with a sex-linked recessive mode of inheritance. These genotypes are shown below.



Individuals 3 and 5 remain ambiguous.

Hence, this pedigree is also consistent with sex-linked recessive inheritance. To evaluate which mode is more likely, we once again count the number of unrelated carriers needed to make this mode of inheritance possible.
Individual:

1) Does not have an allele associated with the trait.
2) Brought her X^n in to the family. Unrelated carrier #1.
3) Does not necessarily have an allele associated with the trait.
4) Got his X^n from 2.
5) If she has an X^n, she got it from 2.
6) Got her X^n from 2.
7) Does not have an allele associated with the trait.
8) Does not have an allele associated with the trait.
9) Does not have an allele associated with the trait.
10) Got his X^n from 6.
11) Does not have an allele associated with the trait.

Therefore, a sex-linked recessive mode of inheritance requires only one unrelated carrier, whereas autosomal recessive inheritance required three unrelated carriers. If the trait is rare, then unrelated carriers are rare, so sex-linked recessive is a more likely mode of inheritance.

iii) Individuals 6 and 7 are X^NX^n and X^NY. The Punnett square is:

	X^N	X^n
X^N	X^NX^N	X^NX^n
Y	X^NY	X^nY

← sons

The sons will be 1/2 normal and 1/2 affected. Therefore, the chance that the next son will be affected is 1/2 or 50%.

(1.5.2)
a) First eliminate modes of inheritance that are not possible. Autosomal dominant inheritance is impossible because affected individuals 7, 8, and 13 have parents that are all normal.

Both autosomal recessive and sex-linked recessive modes are possible. However, a sex-linked recessive mode of inheritance is more likely because 1) there are five affected males and only one affected female, and 2) a sex-linked recessive mode requires only two unrelated carriers, individuals 1 and 2. (See solutions to problem [1.5.1] for more information on unrelated carriers.) Although this pedigree is consistent with inheritance of an autosomal recessive trait, this would require five unrelated carriers (1, 2, 3, 6, and 11).

b)

Genotype	Phenotype
X^NX^N	normal female
X^NX^n	normal female
X^nX^n	affected female
X^NY	normal male
X^nY	affected male

1: X^nY 2: X^NX^n 3: X^NY 4: X^NX^n 5: X^nX^n 6: X^NY 7: X^nY
8: X^nY 9: X^nY 10: X^NX^n 11:X^NY 12: X^NY 13: X^nY

(1.5.3)
a) First eliminate modes of inheritance that are not possible.

- not autosomal recessive: 4 and 5 are affected but they have a normal child (7).
- not sex-linked recessive: for two reasons:
  - 4 and 5 are affected but they have a normal child (7)
  - 1 is affected but she has a normal son (3)

Therefore, an autosomal dominant mode of inheritance is likely. It is important to work through the pedigree to make sure it is consistent.

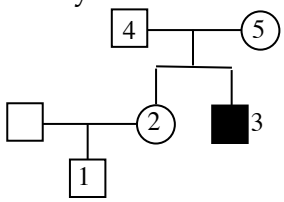
b)

Genotype	Phenotype
DD or Dd	affected (dominant)
dd	normal (recessive)

c) 1: Dd 2: dd 3: dd 4: Dd 5: Dd 6: Dd 7: dd

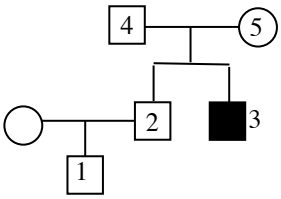
d) Dd × Dd gives 3:1 affected: normal. Therefore, there is a 75% chance that she will be affected.

(1.5.4) The pedigree for "a male those whose mother's brother is a hemophiliac" is represented below. Why should individual 1 be exempt from circumcision?



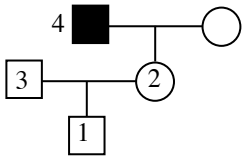
a) Individual 3 has hemophilia, although his parents do not have the disease. Thus, individual 5 must be a carrier (X^hX^H). Individual 3 would have received an X^h allele from his mother (individual 5) and a Y from his father (individual 4). Therefore, individual 2 has a 50% chance of being a carrier and if she is a carrier, the son (individual 1) will have a 50% chance of being a hemophiliac. This means that a male whose mother's brother is a hemophiliac has a 25% chance of being a hemophiliac. Exemption from circumcision makes sense.

b) This is not an oversight. See the pedigree below.



Again, individual 3 has hemophilia, although his parents do not have the disease. Thus, individual 5 must be a carrier (X^hX^H). However, individual 2 does not have the disease so he must have received the X^H allele from his mother. Individual 2 has the genotype X^HY, he does not carry the disease allele and cannot then pass it to the child. Therefore, individual 1 has no more risk for hemophilia than anyone in the general population.

c) Should an exemption be made for the son of a mother whose father is a bleeder? Explain.



Individual 4 has hemophilia. Thus, individual 2 would have received an X^h allele from her father and must be a carrier (X^hX^H). Since individual 2 is a carrier, individual 1 has a 50% chance of being a hemophiliac. An exemption from circumcision makes sense.

(1.5.5) We will use symbols outlined in earlier problems.

For autosomal recessive:

Genotype	Phenotype
NN or Nn	normal
nn	diseased

For sex-linked recessive:

Genotype	Phenotype
X^NX^N	normal female
X^NX^n	normal female
X^nX^n	diseased female
X^NY	normal male
X^nY	diseased male

For autosomal dominant:

Genotype	Phenotype
DD or Dd	diseased (dominant)
dd	normal (recessive)

a) First eliminate modes of inheritance that are not possible.

- A diseased mother has a normal son, so sex-linked recessive is not possible.
- Autosomal recessive is possible, but it requires four unrelated carriers.

  i) So an autosomal dominant mode of inheritance is more likely with only one unrelated carrier.

  ii) * = Dd

b) First eliminate modes of inheritance that are not possible.

- Two normal parents have a diseased child, so autosomal dominant is not possible.
- A diseased mother has a normal son, so sex-linked recessive is not possible.

  i) An autosomal recessive mode of inheritance is consistent with this pedigree.

  ii) * = Nn (not NN because the father is nn)

c) First eliminate modes of inheritance that are not possible.

- Two normal parents have a diseased child, so autosomal dominant is not possible.
- Autosomal recessive is possible, but it requires four unrelated carriers.
- Sex-linked recessive requires only two unrelated carriers.

  i) A sex-linked recessive mode of inheritance is more likely.
  ii) * = X^NX^n

d) First eliminate modes of inheritance that are not possible.

- Two normal parents have a diseased child, so autosomal dominant is not possible.
- A normal father has a diseased daughter, so sex-linked recessive is not possible.

i) Autosomal recessive is the only possible mode.
ii) * = Nn

e) First eliminate modes of inheritance that are not possible.

- Two normal parents have a diseased child, so autosomal dominant is not possible.
- Autosomal recessive is possible but requires four unrelated carriers.
- Sex-linked recessive requires only one unrelated carrier.

i) Sex-linked recessive
ii) * = X^nY

f) First eliminate modes of inheritance that are not possible.

- Two normal parents have a diseased child, so autosomal dominant is not possible.
- Autosomal recessive is possible but requires four unrelated carriers.
- Sex-linked recessive requires only one unrelated carrier.

i) Sex-linked recessive
ii) * = X^NX^n

g) All three modes are possible.

- Autosomal recessive is possible but requires three unrelated carriers
- Sex-linked recessive is possible but requires three unrelated carriers
- Autosomal dominant requires only one unrelated carrier.

i) Autosomal dominant
ii) * = dd

(1.6) One gene; more complex models, II

(V5) Virtual Genetics Lab V
There are no solutions for this part.

(V6) Virtual Genetics Lab VI
There are no solutions for this part.

(2) PROBLEMS INVOLVING TWO OR MORE GENES

(2.1) Two or more genes that assort independently (Mendel's Second Law)

(2.1.1)

a) *Green tall (GGTT)* × *red short (ggtt)*.

As before, the rows and columns in the Punnett square correspond to *gametes* made by each parent. Each gamete has one copy of each gene. Therefore, GT is a possible gamete (it has one copy of the color gene and one copy of the height gene), but GG is not (it has two copies of the color gene and no copies of the height gene). Both parents produce only one kind of gamete because they are homozygous for both the color gene and the height gene. For example, the green tall parent has two copies of the G allele and two copies of the T allele ($G_1G_2T_1T_2$) and the possible gametes are G_1T_1, G_1T_2, G_2T_1, and G_2T_2. Thus, the Punnett square is:

	G_1T_1 or GT	G_1T_2 or GT	G_2T_1 or GT	G_2T_2 or GT
gt	GgTt	GgTt	GgTt	GgTt
gt	GgTt	GgTt	GgTt	GgTt
gt	GgTt	GgTt	GgTt	GgTt
gt	GgTt	GgTt	GgTt	GgTt

All have the genotype GgTt and are green and tall.

b) *Green tall (GgTt)* × *red short (ggtt)*.

The Punnett square is:

	GT	Gt	gT	gt
gt	GgTt green, tall	Ggtt green, short	ggTt red, tall	ggtt red, short
gt	GgTt green, tall	Ggtt green, short	ggTt red, tall	ggtt red, short
gt	GgTt green, tall	Ggtt green, short	ggTt red, tall	ggtt red, short
gt	GgTt green, tall	Ggtt green, short	ggTt red, tall	ggtt red, short

This gives the following offspring:

1/4 GgTt green tall
1/4 Ggtt green short
1/4 ggTt red tall
1/4 ggtt red short

Note that the GgTt parent produces four different kinds of gametes because it has two alleles of the color gene to choose from and two alleles of the height gene to choose from when making gametes.

c) *Green short (Ggtt)* × *red tall (ggTt)*.

The Punnett square is:

	Gt	Gt	gt	gt
gT	GgTt	GgTt	ggTt	ggTt
gT	GgTt	GgTt	ggTt	ggTt
gt	Ggtt	Ggtt	ggtt	ggtt
gt	Ggtt	Ggtt	ggtt	ggtt

This gives the following offspring:
1/4 GgTt green tall
1/4 Ggtt green short
1/4 ggTt red tall
1/4 ggtt red short

d) *Green short (GGtt)* × *Green tall (GgTt)*.

The Punnett square is:

	Gt	Gt	Gt	Gt
GT	GGTt	GGTt	GGTt	GGTt
Gt	GGtt	GGtt	GGtt	GGtt
gT	GgTt	GgTt	GgTt	GgTt
gt	Ggtt	Ggtt	Ggtt	Ggtt

This gives the following offspring:
1/2 G_Tt green tall
1/2 G_tt green short

e) *Green tall (GgTt)* × *Green tall (GgTt)*.

The Punnett square is:

	GT	Gt	gT	gt
GT	GGTT	GGTt	GgTT	GgTt
Gt	GGTt	GGtt	GgTt	Ggtt
gT	GgTT	GgTt	ggTT	ggTt
gt	GgTt	Ggtt	ggTt	ggtt

This gives the following offspring:
9/16 G_T_ green tall
3/16 G_tt green short
3/16 ggT_ red tall
1/16 ggtt red short

(2.1.2)
a) One approach to solving these problems is to consider each characteristic separately:
• Eye color:
The first cross simplifies to: red × red gives 75% red and 25% white. The ratio is 3:1, which is what you expect if red is dominant to white and the parents are heterozygotes.

Genotype	Phenotype
RR or Rr	red eyes (dominant)
rr	white eyes (recessive)

• Wing shape:
The first cross simplifies to: curly × straight gives all curly. This is what you expect if curly is dominant to straight and the parents are both homozygotes.

Genotype	Phenotype
CC or Cc	curly wings (dominant)
cc	straight wings (recessive)

• Body shape:
The first cross simplifies to: normal × tubby gives all tubby. This is what you expect if tubby is dominant to normal and the parents are both homozygotes.

Genotype	Phenotype
TT or Tt	tubby body (dominant)
tt	normal body (recessive)

b) The genotypes of the parents are therefore:
RrCCtt (red-eyed, curly-winged, normal body)
RrccTT (red-eyed, straight-winged, tubby body)

c) The genotype of the white-eyed, curly-winged, tubby F_1's will be: rrCcTt
So the cross is rrCcTt × rrCcTt, which gives:

9	rrC_T_	white-eyed, curly-winged, tubby body
3	rrC_tt	white-eyed, curly-winged, normal body
3	rrccT_	white-eyed, straight-winged, tubby body
1	rrcctt	white-eyed, straight-winged, normal body

(2.1.3)
a) Examine one characteristic at a time:
• Eye color:
Cross 1: red × red gives 42 red
Cross 2: black × black gives 32 black and 11 red
Cross 3: black × black gives 41 black
Cross 4: black × red gives 32 black
Crosses 4 and 2 suggest that black is the dominant phenotype.

• Body color:
Cross 1: brown × brown gives 42 brown
Cross 2: green × green gives 43 green
Cross 3: green × green gives 31 green and 10 brown
Cross 4: green × brown gives 32 green
Crosses 3 and 4 suggest that green is the dominant phenotype.

b) Eye color:

Genotype	Phenotype
BB or Bb	black eyes (dominant)
bb	red eyes (recessive)

Body color:

Genotype	Phenotype
GG or Gg	green body (dominant)
gg	brown body (recessive)

c) Cross 1: bbgg × bbgg gave: all bbgg red eyes, brown body

Cross 2: BbG_ × BbGG (or BbGG × BbG_)
gave: 3 B_G_ black eyes, green body
1 bbG_ red eyes, green body

Cross 3: B_Gg × BBGg (or BBGg × B_Gg)
gave: 3 B_G_ black eyes, green body
1 B_gg black eyes, brown body

Cross 4: B_GG × BBgg (or BBGG × B_gg)
gave: all B_Gg black eyes, green body

(2.1.4)
a) The F_1 plants, when crossed with each other, give progeny of four phenotypic classes in a 9:3:3:1 ratio, suggesting that the F_1 plants are heterozygous at both the leaf shape gene and the stem height genes. In a cross of two individuals that are heterozygous for two genes, the predominant class of progeny will be the one showing the dominant phenotype for both traits. The predominant phenotype in the F_2 generation is broad leaves and long stems; thus, broad leaves and long stems are the dominant phenotypes.

b) As stated in (a), the F_1 plants are heterozygous for both genes. Some appropriate symbols are:

Genotype	Phenotype
BB or Bb	broad leaves (dominant)
bb	narrow leaves (recessive)

Genotype	Phenotype
LL or Ll	long stem (dominant)
ll	short stem (recessive)

The genotype of the F_1 plants is therefore BbLl.

c) Normally in a simple two-factor cross, we expect to see four phenotypes in the F_2 generation, resulting from all possible combinations of leaf shape and stem length. In the F_2 generation of this cross, however, plants with narrow leaves and long stems are missing and presumed dead. One plausible explanation for this missing phenotype is that seedlings of the genotype bbL_ die because the narrow leaves of this plant cannot supply the energy demands of a long-stemmed adult plant.

d) Since the genotype of the F_1 plants is BbLl (see part b), the two parental plants could have had genotypes of BBLL and bbll or BBll and bbLL. However, we know that any bbLL plant dies before it reaches adulthood (see part c). Thus, one parent must have had broad leaves and a long stem (BBLL) and the other had narrow leaves and a short stem (bbll).

(2.2) Two or more genes that are linked

(2.2.1) An organism is true breeding when, for several generations, the phenotype of the parents is the only phenotype seen in the offspring. For example, if you cross two animals that have black fur and long ears then you would see only offspring with black fur and long ears. This occurs when the parents are homozygous.

a) The true-breeding animals with black fur and long ears are BBLL. The true-breeding animals with red fur and short ears are bbll. We have been told that these two loci are linked so the genotype can be shown as: $\frac{BL}{BL}$ and $\frac{bl}{bl}$. The F_1 offspring will get one chromosome from each parent and will be $\frac{BL}{bl}$ (or using symbols that do not indicate linkage, BbLl).

b) If no recombination occurs, then only parental gametes will be formed. Parental or nonrecombinant gametes have the same configuration of alleles as that found in the parent. The configuration of the alleles in the F_1 was $\frac{BL}{bl}$. So half of the gametes produced will be BL and half will be bl.

c) If the genes are linked, but recombination occurs, then more parental gametes than nonparental gametes (BL or bl) will be formed. In nonparental gametes, the configuration of the alleles is different from that seen in the parent. The nonparental (or recombinant) gametes are bL and Bl. A recombination frequency of 10% means that 10 out of every 100 gametes are nonparental. Therefore, the gametes will be:

45% BL
45% bl
5% Bl
5% bL

d) The F_1 was $\frac{BL}{bl}$. The red-furred, short-eared parent is $\frac{bl}{bl}$ and makes only bl gametes, so the resulting offspring would be (roughly):

45	$\frac{BL}{bl}$	or	BbLl	black-furred, long-eared (parental type)
45	$\frac{bl}{bl}$	or	bbll	red-furred, short-eared (parental type)
5	$\frac{Bl}{bl}$	or	Bbll	black-furred, short-eared (nonparental type)
5	$\frac{bL}{bl}$	or	bbLl	red-furred, long-eared (nonparental type)

e) The true-breeding animals with black fur and short ears are BBll. The true-breeding animals with red fur and long ears are bbLL. We have been told that these two loci are linked so the genotype can be shown as: $\frac{Bl}{Bl}$ and $\frac{bL}{bL}$. The F_1 offspring will get one chromosome from each parent and will be $\frac{Bl}{bL}$, all black with long ears.
Note that if you use symbols that do not indicate linkage, this genotype (BbLl) is indistinguishable from that of the F_1 from part (a).

f) The F_1 offspring from (e) are $\frac{Bl}{bL}$. These genes are linked, so more parental gametes (Bl or bL) than nonparental gametes will be formed. The nonparental (or recombinant) gametes are BL or bl. A recombination frequency of 10% means that 10 of every 100 gametes are nonparental. Therefore, the gametes will be:

45%	Bl
45%	bL
5%	BL
5%	bl

Remember that the genotype of the F_1 in part (e) is the same as the genotype of the F_1 from part (a) if you use symbols that do not indicate linkage. These two different F_1 animals produce the same kinds of gametes, but due to the configuration of the alleles, the percentage of each kind of gamete is different. Compare your answers to parts (c) and (f).

g) The F_1 was $\frac{Bl}{bL}$. The red-furred, short-eared parent is $\frac{bl}{bl}$ and makes only bl gametes, so the resulting offspring would be (roughly):

45	$\frac{Bl}{bl}$	or	Bbll	black-furred, short-eared (parental type)
45	$\frac{bL}{bl}$	or	bbLl	red-furred, long-eared (parental type)
5	$\frac{BL}{bl}$	or	BbLl	black-furred, long-eared (nonparental type)
5	$\frac{bl}{bl}$	or	bbll	red-furred, short-eared (nonparental type)

(2.2.2)
a) In both crosses 1 and 2, parents with different phenotypes give rise to offspring that resemble only one parent (black × white ⇒all black or five-toed × six-toed ⇒ all six-toed). Thus, the **dominant** phenotypes are **black** and **six-toed** and the **recessive** phenotypes are **white** and **five-toed.**

Some appropriate symbols are:

Genotype	Phenotype
BB or Bb	black (dominant)
bb	white (recessive)

Genotype	Phenotype
TT or Tt	six-toed (dominant)
tt	five-toed (recessive)

b) This is a test cross. The F_1 individuals produced in cross 2 will be heterozygous for both genes and their genotype will be BbTt. The genotype of the white, five-toed mouse is bbtt.

BbTt × bbtt will give:

1 : BbTt	black, six-toed
1 : Bbtt	black, five-toed
1 : bbTt	white, six-toed
1 : bbtt	white, five-toed

Assuming that the genes for coat color and number of toes are unlinked.

c) Because the ratio of offspring seen varies from that predicted in part (b), there is the possibility that the gene for coat color and the gene for toe number are linked. However, with only seven offspring it is difficult to make a conclusion. It is not possible that white and six-toed and black and five-toed are nonviable combinations, because these combinations have already been observed in cross 1 and cross 2.

d) With additional data, the conclusion can be made that the gene for coat color and the gene for toe number are linked. If we were asked to calculate a recombination frequency for these the two genes, we could make the following calculation:

$$\text{Recombination frequency} = \frac{\#\text{ recombinant progeny}}{\text{total progeny}}$$

In this case #recombinant progeny = 3 + 5 = 8
#total progeny = 57 + 52 + 3 + 5 = 117
Recombination frequency = 0.0684 = 6.84%

(2.2.3)
a) Some appropriate symbols are:

Genotype	Phenotype
RR or Rr	red (dominant)
rr	yellow (recessive)

Genotype	Phenotype
SS or Ss	smooth wings (dominant)
ss	crinkled wings (recessive)

A cross between a true-breeding, yellow-bodied, and smooth-winged female (rrSS) and a true-breeding, red-bodied, and crinkle-winged male (RRss) will give offspring showing both of the dominant phenotypes. The flies will all be red-bodied and smooth-winged (RrSs).

b) The F_1 flies will have the genotype RrSs. In a cross between two F_1 flies, you expect a 9:3:3:1 ratio of phenotypes in the offspring. The data given do not conform to this prediction, so it is possible that the gene for color and the gene for wing type are linked.

Note that if these genes are linked, you would predict more parental type than non-parental type flies. The original cross can be symbolized as:

$$\frac{r\ S}{r\ S}\ \text{female} \times \frac{R\ s}{R\ s}\ \text{male} \longrightarrow F_1 : \frac{r\ S}{R\ s}$$

Therefore, offspring with a parental genotype and phenotype are:

$\frac{rS}{rS}$ Yellow and smooth-winged

$\frac{rS}{Rs}$ Red and smooth-winged

$\frac{Rs}{Rs}$ Red and crinkled-winged

Offspring with a nonparental genotype and phenotype are:

$\frac{RS}{RS}$ Red and smooth-winged

$\frac{RS}{rs}$ Red and smooth-winged

$\frac{rs}{rs}$ Yellow and crinkled-winged

As you can see, there are flies that are red-bodied and smooth-winged in both the parental and nonparental classes. This makes it difficult to evaluate the possible linkage between the gene for color and the gene for wing type. Using a test cross to evaluate the possible linkage between the gene for color and the gene for wing type avoids this problem. See part (c) below.

c) The test cross can be symbolized as:

$$\frac{Rs}{rS} \times \frac{rs}{rs}$$

Which would result in the following offspring:

$\frac{Rs}{rs}$ or Rrss Red, crinkled-winged (parental type)

$\frac{rS}{rs}$ or rrSs Yellow, smooth-winged (parental type)

$\frac{RS}{rs}$ or RrSs Red, smooth-winged (nonparental type)

$\frac{rs}{rs}$ or rrss Yellow, crinkled-winged (nonparental type)

To calculate map units, begin with:

$$\text{Recombination frequency} = \frac{\#\text{ of recombinant progeny}}{\text{total progeny}}$$

$= (102 + 98)/(102 + 98 + 396 + 404) = 200/1{,}000 =$
0.2 or 20% recombination frequency or 20 map units.

d) These two genes are linked.

e) The test cross can be symbolized as:

$$\frac{Rl}{rL} \times \frac{rl}{rl}$$

Which would result in the following offspring:

$\frac{Rl}{rl}$ or Rrll Red, short-winged (parental type)

$\frac{rL}{rl}$ or rrLl Yellow, long-winged (parental type)

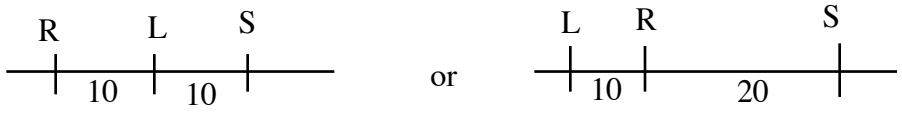
$\frac{RL}{rl}$ or RrLl Red, long-winged (nonparental type)

$\frac{rl}{rl}$ or rrll Yellow, short-winged (nonparental type)

Map distance = 45 + 55/ 45 + 55 + 460 + 440
= 100/1,000 = 0.1, a recombination frequency of 10%
= 10 map units

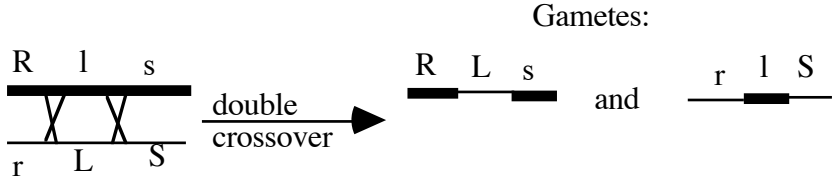
f) Since "r" and "s" are linked, and "r" and "l" are linked, then "s" and "l" must be linked as well.
All three genes are on the same chromosome.

g)



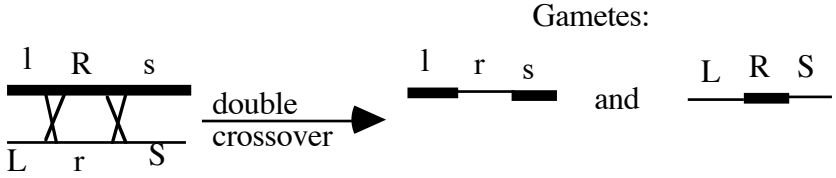
h) $\frac{Rsl}{rSL}$. They are all red-bodied, smooth, and long-winged.

i) Given the genotypes of the parents, if the order were R-L-S, then the double crossover class would have been:

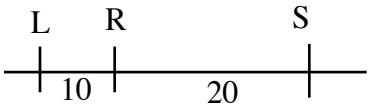


Which would make red/long/crinkled and yellow/short/smooth the two rarest classes. This is not seen; therefore, this order is incorrect.

Given the genotypes of the parents, if the order were L-R-S, then the double crossover class would have been:



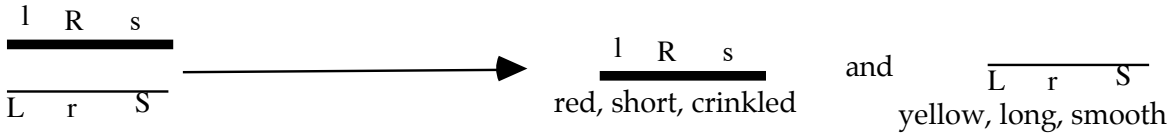
Which would make red/long/smooth and yellow/short/crinkled the two rarest classes. This is observed; therefore, the map is:



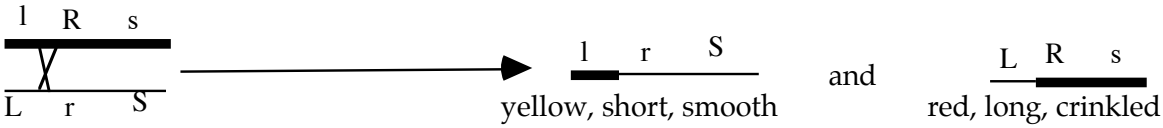
j)

Gametes:

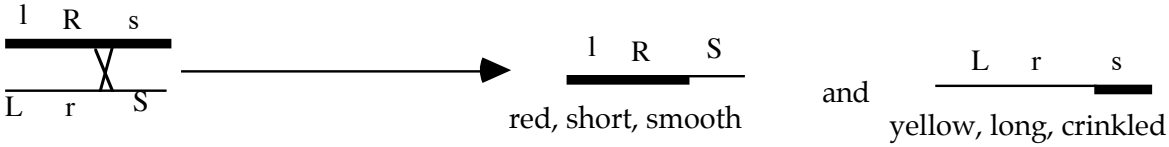
No recombination, PARENTAL types:



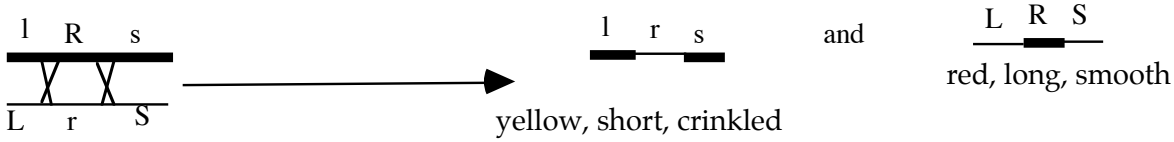
Recombination between L & R:



Recombination between R & S:



Recombination in both regions:



(3) CHALLENGE PROBLEMS

(3.1) When approaching a problem where you are following multiple traits (in this case color and leaf shape) you can begin by separating the traits and evaluating them one at a time. To determine which phenotypes are dominant, first look at the crosses that gave the simplest results.

For example, cross 3 mates a purple plant with a green plant and results in only purple offspring. In this case, the offspring look like one of the parents, so you are dealing with simple dominant inheritance. The lack of green offspring suggests that purple is dominant to green.

Cross 4 mates a plant with sharp leaves to a plant with rounded leaves. In this case, the offspring look like one of the parents, so you are dealing with simple dominant inheritance. The lack of rounded leaves among the offspring suggests that sharp leaves are dominant to rounded leaves

Some appropriate symbols are:

Genotype	Phenotype
PP or Pp	purple (dominant)
pp	green (recessive)

Genotype	Phenotype
CC or Cc	sharp leaves (dominant)
cc	rounded leaves (recessive)

Since the traits segregate independently, they can be treated separately.

For cross 1: purple, sharp × green, sharp

Gives:

Purple sharp	Purple rounded	Green sharp	Green rounded
321	101	310	107

- First evaluate the parents with respect to the color phenotype:

Purple	
321	101

Green	
310	107

or

Purple	Green
422	417

This can be seen as a 1:1 ratio of purple to green. The simplest explanation for this ratio is if one parent were Pp and the other was pp.

- Next evaluate the parents with respect to the leaf phenotype:

Sharp	
321	310

Rounded	
101	107

or

Sharp	Rounded
631	208

This can be seen as a 3:1 ratio of sharp to rounded. The simplest explanation for this ratio is if one parent were Cc and the other was Cc.

a) Therefore, we predict the parents for cross 1 to be PpCc × ppCc.

For cross 2: purple, sharp × purple, rounded

Gives:

Purple sharp	Purple rounded	Green sharp	Green rounded
219	207	64	71

- First evaluate the parents with respect to the color phenotype:

Purple	
219	207

Green	
64	71

or

Purple	Green
426	135

This can be seen as a 3:1 ratio of purple to green. The simplest explanation for this ratio is if one parent were Pp and the other were Pp.

- Next evaluate the parents with respect to the leaf phenotype:

Sharp	
219	64

Rounded	
207	71

or

Sharp	Rounded
283	278

This can be seen as a 1:1 ratio of sharp to rounded. The simplest explanation for this ratio is if one parent were Cc and the other was cc. Also, we know from the cross that one of the parents had rounded leaves and was thus cc.

b) Therefore, we predict the parents for cross 2 to be PpCc × Ppcc.

For cross 3: purple, sharp × green, sharp

Gives:

Purple sharp	Purple rounded	Green sharp	Green rounded
722	231	0	0

- First evaluate the parents with respect to the color phenotype:

Purple	
722	231

Green	
0	0

or

Purple	Green
953	0

All offspring are purple. This predicts that one parent must be PP. The numbers give us no information on the other parent, but we know from the cross that it was green; thus, the genotype of the second parent is pp.

- Next evaluate the parents with respect to the leaf phenotype:

Sharp	
722	0

Rounded	
231	0

This can be seen as a 3:1 ratio of sharp to rounded. The simplest explanation for this ratio is if one parent were Cc and the other was Cc.

c) Therefore, we predict the parents for cross 3 to be PPCc × ppCc.

For cross 4: purple, sharp × green, rounded

Gives:

Purple sharp	Purple rounded	Green sharp	Green rounded
404	0	387	0

- First evaluate the parents with respect to the color phenotype:

Purple	
404	0

Green	
387	0

This can be seen as a 1:1 ratio of purple to green. The simplest explanation for this ratio is if one parent were Pp and the other was pp. Also, we know from the cross that one of the parents was green and thus pp.

- Next evaluate the parents with respect to the leaf phenotype:

Sharp	
404	387

Rounded	
0	0

or all sharp

This predicts that one parent must be CC. We know from the cross that the second parent had rounded leaves and was thus cc.

d) Therefore, we predict the parents for cross 4 to be PpCC × ppcc.

For cross 5: purple, rounded × green, sharp

Gives:

Purple sharp	Purple rounded	Green sharp	Green rounded
70	91	86	77

- First evaluate the parents with respect to the color phenotype:

Purple	
70	91

Green	
86	77

This can be seen as a 1:1 ratio of purple to green. The simplest explanation for this ratio is if one parent were Pp and the other was pp. Also, we know from the cross that one of the parents was green and thus pp.

- Next evaluate the parents with respect to the leaf phenotype:

Sharp	
70	86

Rounded	
91	77

This can be seen as a 1:1 ratio of sharp to rounded. The simplest explanation for this ratio is if one parent were Cc and the other was cc. Also, we know from the cross that a parent had rounded leaves and was thus cc.

e) Therefore, we predict the parents for cross 5 to be Ppcc × ppCc.

(3.2)
To begin a problem like this, look at each phenotype separately.

For wing shape, you see roughly 1:1 bent to straight phenotypes and both males and females of each. This predicts that one of the parents is a homozygote and the other is a heterozygote. From the F_1, we cannot tell which phenotype is dominant. However, in the second mating, we see a 3:1 ratio of bent to straight. Thus, we conclude that bent is dominant to straight.

Some appropriate symbols would be:

Genotype	Phenotype
BB or Bb	bent wings (dominant)
bb	straight wings (recessive)

In the first cross: a straight-winged, yellow-eyed female × a bent-winged, red-eyed male gave:

41 bent, yellow males
38 bent, red females
38 straight, yellow males
43 straight, red females

You see roughly 1:1 yellow to red phenotypes. However, upon closer examination, the yellow phenotype is seen only in males. This suggests that eye color is sex linked, and the red-eyed phenotype is dominant to the yellow-eyed phenotype. Another possibility is that yellow-eyed females are inviable; however, the female in the cross is yellow-eyed, so you assume eye color is a sex-linked trait.
Some appropriate symbols would be:

Genotype	Phenotype
X^RX^R or X^RX^r	female with red eyes (dominant)
X^rX^r	female with yellow eyes (recessive)
X^RY	male with red eyes (dominant)
X^rY	male with yellow eyes (recessive)

Therefore, the parents in the first cross would have been:
a straight-winged, yellow-eyed female (bbX^rX^r) × a bent-winged, red-eyed male (BbX^RY)

In the second cross we see both wing types in the progeny at a 3:1 ratio of bent to straight. This suggests that both the male and female F_1 flies are heterozygotes (Bb). In the second cross, both red-eyed and yellow-eyed males are produced, so the F_1 red-eyed female must be heterozygous for eye color, (X^RX^r). Last, we know the F_1 yellow-eyed male must be X^rY.

Thus, the second cross would be:
An F_1 bent-winged, red-eyed female (BbX^RX^r) × an F_1 bent-winged, yellow-eyed male (BbX^rY)

(3.3) You must first assume that in the statement "*all* the progeny are tall, pink plants" that *all* is significant because many progeny were evaluated. Then treat each trait independently: a tall plant crossed with a short plant yields all tall plants, which suggests that the tall phenotype is dominant to the short phenotype and that the tall parent was a homozygote with respect to the height gene.
Some appropriate symbols would be:

Genotype	Phenotype
TT or Tt	tall (dominant)
tt	short (recessive)

A white plant crossed with a red plant yields all pink plants. A plausible model is that color is controlled by one gene with two alleles and that color is incompletely dominant such that:

Genotype	Phenotype
CC	white
CC′	pink
C′C′	red

a) The genotype of the tall, white parental plant is TTCC.

b) The genotype of the short, red parental plant is ttC′C′.

c) The genotype of an F_1 tall, pink plant is TtCC′.

d) The six phenotypic classes of plant and associated genotype(s) are:

Class	Phenotype	Genotype
1	tall, white	TTCC and TtCC
2	tall, pink	TTCC′ and TtCC′
3	tall, red	TTC′C′ and TtC′C′
4	short, white	ttCC
5	short, pink	ttCC′
6	short, red	ttC′C′

e) Pure-breeding plants are those homozygous for a trait. In this case the pure-breeding genotypes are TTCC, TTC′C′, ttCC, and ttC′C′. However, there is no way to tell what the genotype of a tall plant is with respect to the height locus. Therefore, the only pure-breeding plants you can distinguish by looking at the phenotype are short, white-flowered plants and short, red-flowered plants.

f) TtCC′ × TtCC′ would yield the following if the two genes are unlinked.

	TC	TC′	tC	tC′
TC	TTCC Tall, white	TTCC′ Tall, pink	TtCC Tall, white	TtCC′ Tall, pink
TC′	TTCC′ Tall, pink	TTC′C′ Tall, red	TtCC′ Tall, pink	TtC′C′ Tall, red
tC	TtCC Tall, white	TtCC′ Tall, pink	ttCC Short, white	ttCC′ Short, pink
tC′	TtCC′ Tall, pink	TtC′C′ Tall, red	ttCC′ Short, pink	ttC′C′ Short, red

(3)	(6)	(3)	(1)	(2)	(1)
Tall, white	Tall, pink	Tall, red	Short, white	Short, pink	Short, red

This ratio deviates from the 9:3:3:1 that you are accustomed to in a di-hybrid cross because of the incomplete dominance of the color phenotype.

g) If the two genes are linked, we could designate the parents as: $\frac{TC}{TC}$ and $\frac{tC'}{tC'}$.

Therefore, the F_1 would be $\frac{TC}{tC'}$.

Because we have been told that there is no recombination, all the offspring from a cross between F_1 plants will get either TC or tC′ from each parent. The Punnett square becomes:

	TC	tC′
TC	TTCC Tall, white	TtCC′ Tall, pink
tC′	TtCC′ Tall, pink	ttC′C′ Short, red

(3.4)

a) Using the law of independent assortment for the two traits, the Punnett square is:

	tX^R	tX^R	tX^r	tX^r
TX^R	Tt X^RX^R Red, female	Tt X^RX^R Red, female	Tt X^RX^r Red, female	Tt X^RX^r Red, female
TY	Tt X^RY Red, male	Tt X^RY Red, male	Tt X^rY White, male	Tt X^rY White, male
tX^R	tt X^RX^R Red, female	tt X^RX^R Red, female	tt X^RX^r Red, female	tt X^RX^r Red, female
tY	tt X^RY Red, female	tt X^RY Red, female	tt X^rY Red, female	tt X^rY Red, female

This gives:

12/16 or (3/4) red female
2/16 or (1/8) red male
2/16 or (1/8) white male

overall 3:1 female : male.

b) The Punnett square would be:

	tX^R	tX^R	tX^r	tX^r
TX^R	Tt X^RX^R Red, female	Tt X^RX^R Red, female	Tt X^RX^r Red, female	Tt X^RX^r Red, female
TY	Tt X^RY Red, male	Tt X^RY Red, male	Tt X^rY White, male	Tt X^rY White, male
tX^R	tt X^RX^R Red, female	tt X^RX^R Red, female	tt X^RX^r Red, female	tt X^RX^r Red, female
tY	tt X^RY Dead	tt X^RY Dead	tt X^rY Dead	tt X^rY Dead

This gives (since only 12/16 of these offspring live, the fractions are out of 12 possible):

8/12 or (2/3) red female
2/12 or (1/6) red male
2/12 or (1/6) white male

overall 5:1 female : male.

(3.5) Cystic fibrosis is an autosomally inherited genetic disease. The disease phenotype is recessive to the normal phenotype.
Some appropriate symbols are:

Genotype	Contribution to phenotype
DD or Dd	normal (dominant)
dd	cystic fibrosis (recessive)

a) A carrier has genotype Dd. Therefore, Dd × Dd have a 1 in 4 (25%) chance of producing a dd (cystic fibrosis phenotype) child.

b) To be unaffected and have an affected child, both parents must be carriers (Dd). Therefore, the risk is the same as part (a): 25%.

c) To have cystic fibrosis, the child must inherit a d allele from each parent. The affected parent is dd and will pass a d on to the child. This couple can produce an affected child only if the unaffected parent is a carrier and passes the d allele on to the child. From the question we know that the odds that the unaffected parent is a carrier (Dd) are 1/1,000. If the unaffected parent is a carrier, the chances of passing on the d allele are 1/2. Therefore, the chance of this couple having an affected child is 1/1,000 × 1/2, or 1/2,000.

(3.6)
Keep in mind that this particular linkage applies to this pedigree only. In different families, you could find the disease allele associated with a different blood type allele.
Some appropriate symbols are:

Blood type:

Genotype	Phenotype
I^AI^A or I^Ai	blood type A
I^BI^B or I^Bi	blood type B
I^AI^B	blood type AB
ii	blood type O

Qase:

Genotype	Phenotype
QQ or Qq	normal
qq	Qosis disease

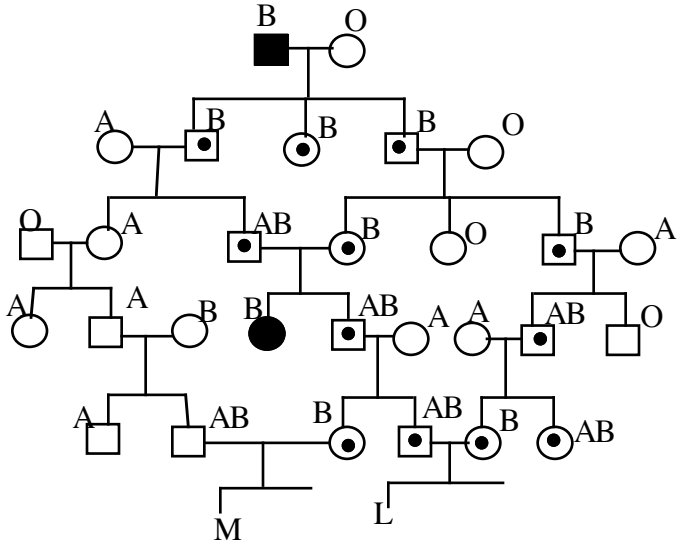
a) In this pedigree, the disease allele is linked to the type B allele. You could assume that the affected male in the first generation is:

$$\frac{I^B \quad q}{I^B \quad q} \text{ or } \frac{I^B \quad q}{i \quad q}$$

and that the second generation of the family would be (reading left to right):

$$\frac{I^A \quad Q}{I^A \quad Q}, \frac{I^B \quad q}{i \quad Q}, \frac{I^B \quad q}{i \quad Q}, \frac{I^B \quad q}{i \quad Q}, \text{ and } \frac{i \quad Q}{i \quad Q}$$

If a dot (•) indicates a carrier for Qosis, then the pedigree would look like this:



b) In this family…

- If L has blood type A, then the risk that child L has Qosis is 0%. If child L has type A blood, then his/her genotype is $\frac{I^A \quad Q}{i \quad Q}$. He/she received the i allele from the carrier mother, but the i allele is not associated with the q disease allele.

- If child L has blood type B, then the risk that child L has Qosis is 50%. If child L has type B blood, then his/her genotype is either $\frac{I^B \quad q}{i \quad Q}$ or $\frac{I^B \quad q}{I^B \quad q}$.

- If child L has blood type AB, then the risk that child L has Qosis is 0%. If child L has type AB blood, then his/her genotype is either $\frac{I^A \quad Q}{I^B \quad q}$ and he/she is a carrier of Qosis.

c) Child M has no risk of having Qosis. Child M will get either I^A Q or I^B Q from dad, so will never show the recessive disease phenotype.

d) If recombination could occur, then the I^B and q alleles would become separated genetically and you could no longer use the blood type test with certainty. Blood type B in this family would no longer indicate a carrier. As the recombination frequency increases, the uncertainty of the diagnosis increases.

(3.7)

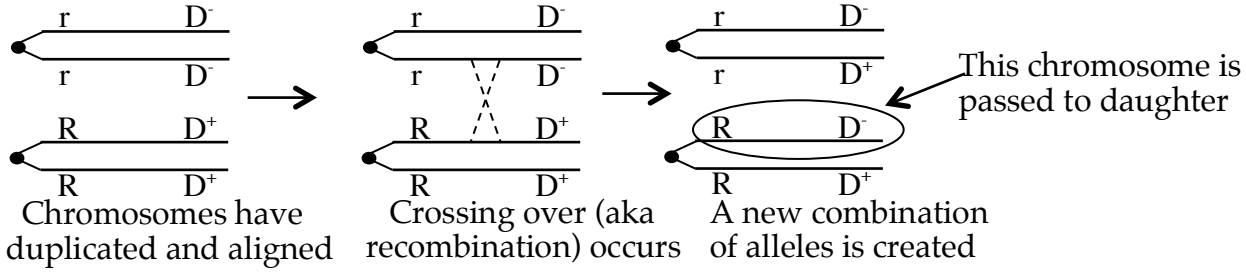
a) The genotypes of individuals 1–5 are as follows:

$1 = \frac{R \quad X^+}{r \quad X^\pm}$ $2 = \frac{R \quad D^+}{r \quad D^\pm}$ $3 = \frac{r \quad D^\pm}{r \quad D^\pm}$ $4 = \frac{R \quad D^+}{r \quad D^\pm}$ $5 = \frac{r \quad D^\pm}{r \quad D^\pm}$

b) The disease allele is linked to the (r) allele since all of the affected individuals but one have blue eyes. This is the $r \quad D^-$ copy of the chromosome.

c) The genotype of this individual is $\frac{R \quad D^-}{r \quad D^-}$; therefore, a crossover event has occurred during meiosis in her mother: $\left(genotype : \frac{R \quad D^+}{r \quad D^-} \right)$.

This schematic shows the crossover event during Prophase I in the gamete-producing cells in the individual marked with a ▲ on the pedigree shown in the problem.



d) This cub gets an $r \quad D^\pm$ chromosome from her father no matter what. In order for the cub to have red eyes, it must have gotten an R from the mother. Since the frequency of recombination is 15%:

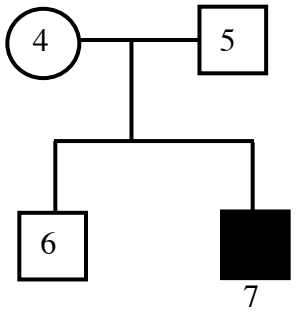
• 85% of the time there will be no recombination between eye color and the disease allele in the mother and the mother will give an $R \quad D^+$ chromosome, resulting in a non-affected pup.

• 15% of the time there will be recombination between eye color and the disease allele in the mother and the mother will give an $R \quad D^-$ chromosome, resulting in an affected pup.

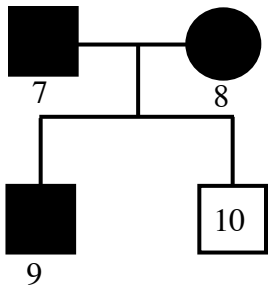
Therefore, the risk that a red pup will have the disease is 15%. Note that this is less than if the genes were unlinked (50%).

(3.8)

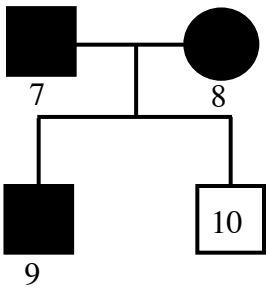
a) For autosomal dominant inheritance, affected children must have at least one affected parent. This part of the pedigree is inconsistent with an autosomal dominant mode of inheritance.



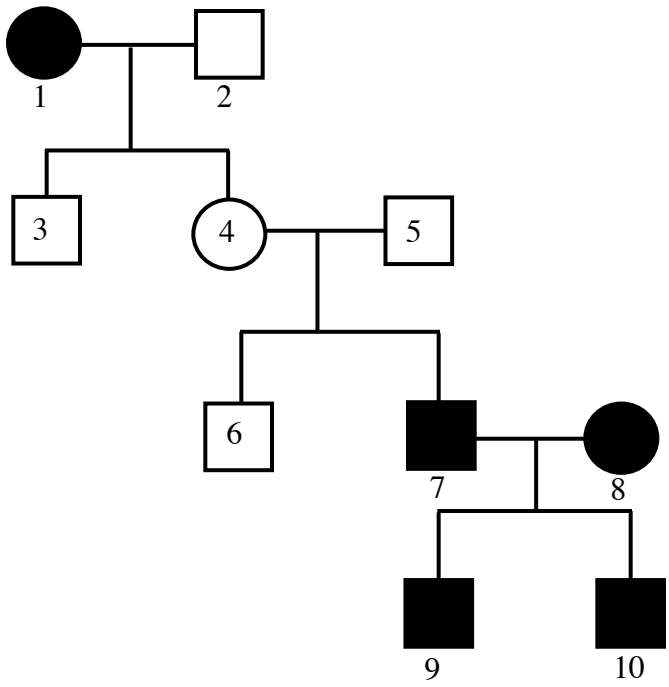
b) For autosomal recessive inheritance, two affected parents will have only affected children. This part of the pedigree is inconsistent with an autosomal recessive mode of inheritance.



c) For sex-linked recessive inheritance, all the sons from an affected mother will be affected. This part of the pedigree is inconsistent with a sex-linked recessive mode of inheritance.



d) If you change 10 from unaffected to affected, the pedigree is now consistent with an autosomal recessive mode of inheritance.



(3.9) In general, start with the simplest model you can think of and test it against the data.

a) There are three possible models to explain the phenotypes of these aliens.

1) One possibility is that blood type is controlled by one gene with the three alleles I^{α}, I^{β}, and I^{γ}. This theory fits only with crosses 4 and 5. The other crosses do not fit this theory.

2) Another way to explain the three phenotypes is to assume that blood type is due to one gene with two alleles, but the heterozygote is a different blood type than either homozygote. In cross 1, the three different phenotypes are present in the offspring, so one of the parents must be the heterozygote. For example:

TT type β
T′T type α
T′T′ type γ

If you apply these genotypes to cross 1, however, it is clear that this theory alone cannot explain the data.

3) If you examine cross 2 (male β × female γ) and cross 3 (male γ × female β), you see different results even though both crosses are β × γ. You also see that blood type is not evenly distributed between the sexes. This suggests that blood type is a sex-linked trait.

- First consider if XX = female and XY = male fit the data. Assume that blood type is due to one gene with two alleles where the heterozygote is a different blood type than either homozygote (see above). If XX = female and XY = male, then a heterozygote female will give male offspring of two different phenotypes, both of which are different from the phenotype seen in the heterozygote XX female.

- Next consider if ZZ = male and ZW = female fit the data. Assume that blood type is due to one gene with two alleles where the heterozygote is a different blood type than either homozygote (see above). If ZW = female and ZZ = male, then a heterozygote male will give female offspring of two different phenotypes, both of which are different from the phenotype seen in the heterozygote ZZ male.

Cross 1 fits the model that ZZ = male and ZW = female and that blood type is due to one gene with two alleles where the heterozygote is a different blood type from either homozygote. This model predicts that blood type is inherited in the following manner:

Z^AZ^A = male γ	Z^AW = female γ
Z^AZ^a = male α	
Z^aZ^a = male β	Z^aW = female β

The data from each cross are consistent with what is predicted by this model.

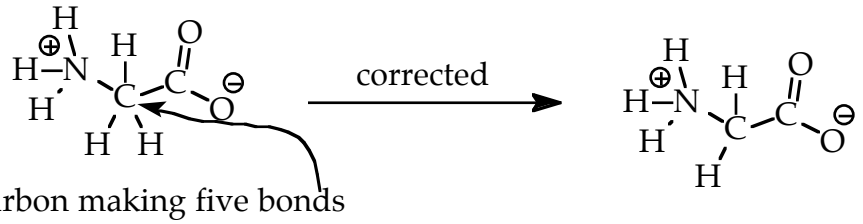
b) In the above model, blood type α is the phenotype of a heterozygote. Also in this model females carry only one sex chromosome (ZW = female). Therefore, it is not possible to have a female with type α blood.

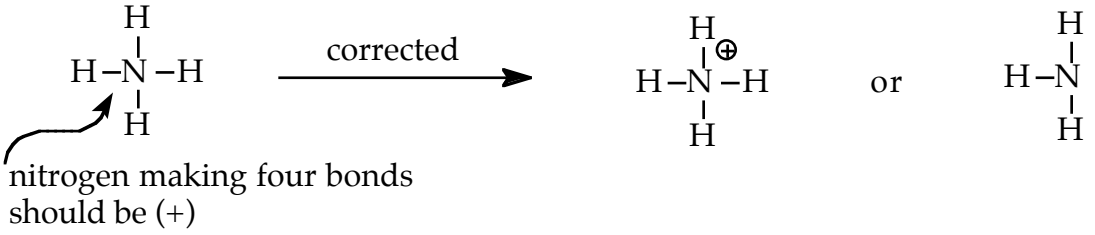
Biochemistry Solutions

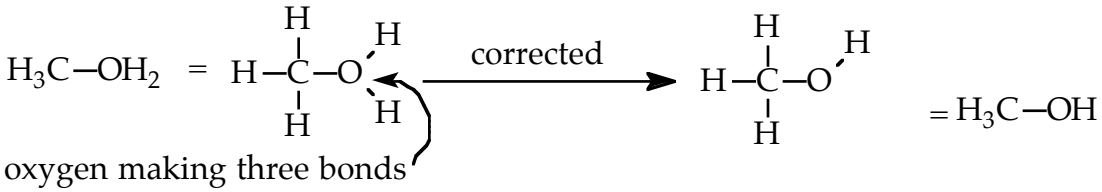
(1) BONDS AND FORCES

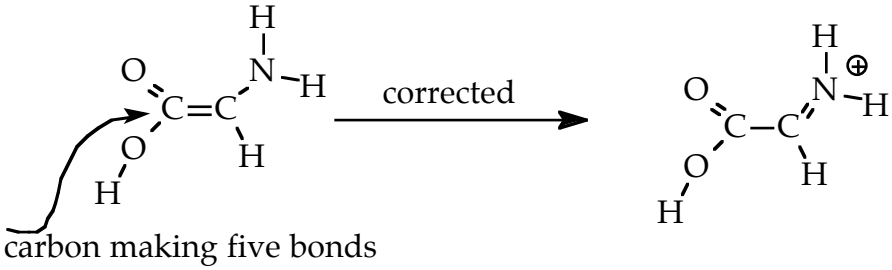
(1.1) Covalent Bonds

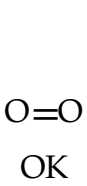
(1.1.1)

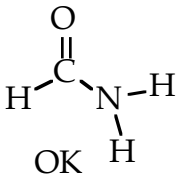










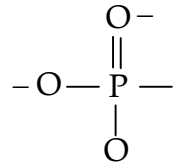


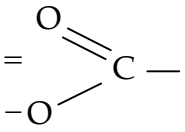
(1.1.2) For each of the functional groups given, draw a structural formula.

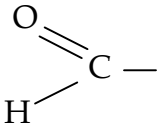
Amino = $^{+}H_3N$ —

Methyl = H_3C —

Hydroxyl = HO —

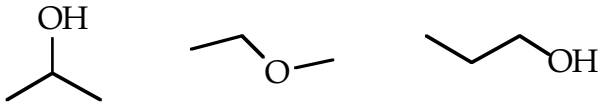
Phosphoryl = 

Carboxyl = 

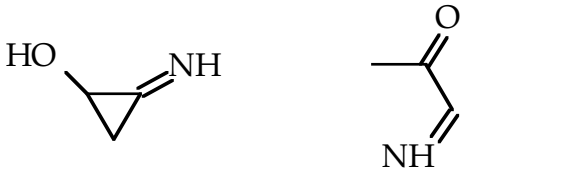
Aldehyde = 

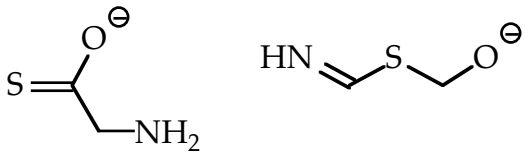
(C1) Computer-Aided Problems 1

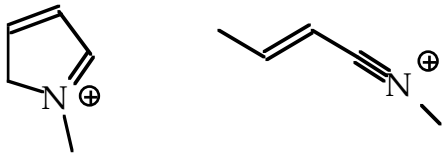
(1.1.3)

a) 

b) 

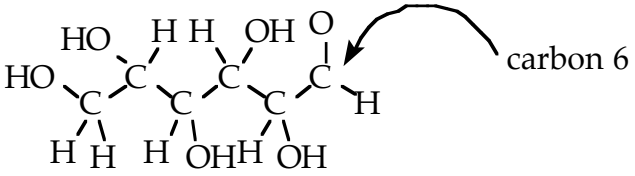
c) 

d) 

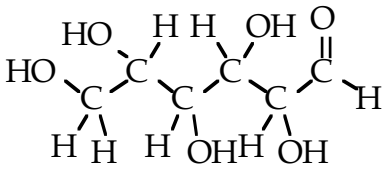
e) 

(C2) Computer-Aided Problems 2

a) The best way to solve these problems is to reproduce what you see on the screen and then go back to fix the single/double/triple bonds. From the screen, you get this:

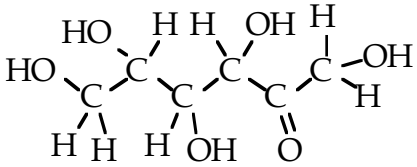


What about carbon 6? The oxygen is shown making only one bond to the carbon. It could be (–) charged or it could be making a double bond with the carbon. If you check the carbon, it is making only three bonds, not four. Therefore, the most reasonable structure is:



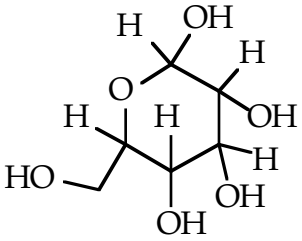
This is the same as the structure shown in most textbooks.

b) Using the same logic as you did with part (a), you get:



This should match the structure in the textbook. Compared with glucose, the C=O is on the second-to-last carbon of fructose but on the last carbon of glucose. The –OHs are also switched.

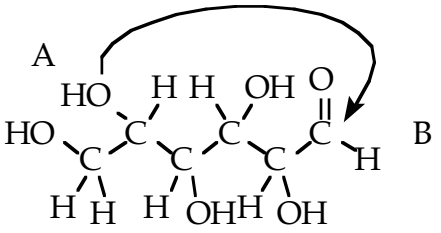
c) The structure looks like this:



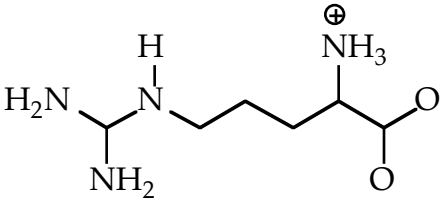
How is this related to the linear form? The key is to compare the carbon atoms. In linear glucose, each carbon has one and only one oxygen atom attached. In the circular form, one carbon atom (the one at the top of the hexagon) has two oxygens attached. This means that one of the oxygen atoms on one carbon is attaching to another carbon atom to form the ring.

This also means that the carbon at the top of the ring must have been at one end of the chain. Therefore, the oxygen marked A in the structure below bonded to the carbon

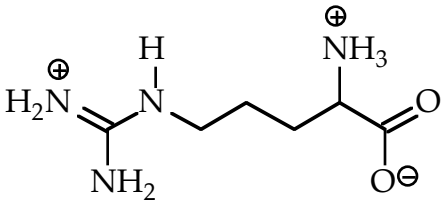
marked B. The H that was attached to A becomes attached to the oxygen that is attached to carbon B. This is shown below:

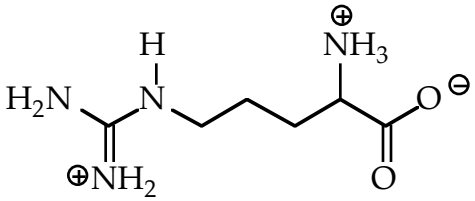


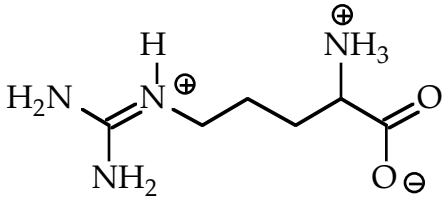
d) The structure shown by the program is:



There are several possible ways the double bonds could be arranged:

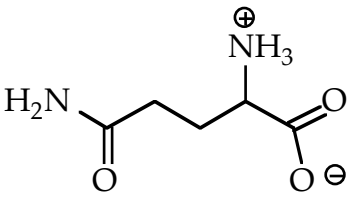




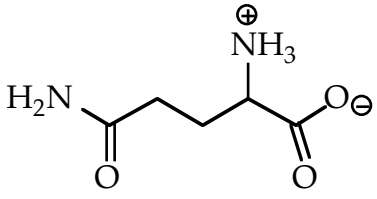


and several other combinations (six total).
In fact, the actual situation is a mixture of all six forms (this is called resonance). For our purposes, it does not matter which one you choose; they are all arginine.

e) The structure is:

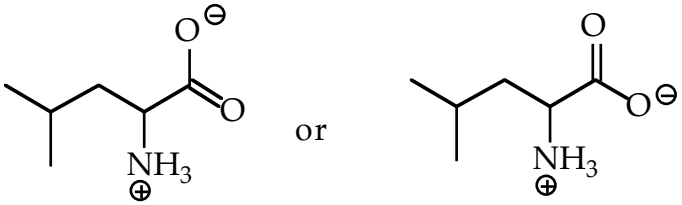


or



Both are glutamine.

f) The structure is:

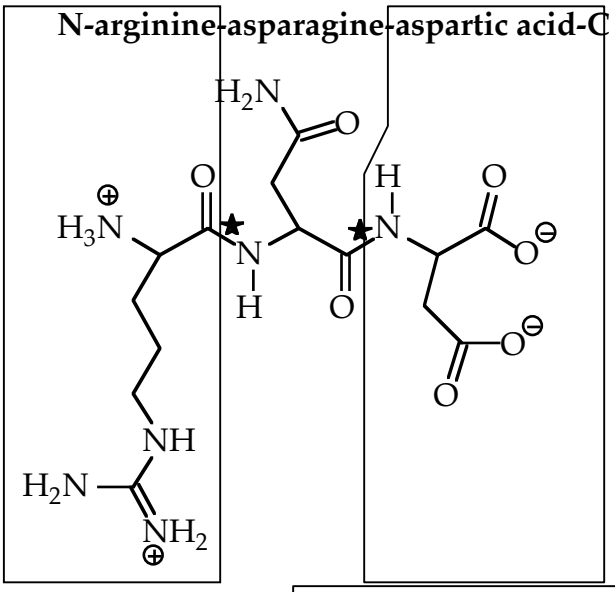


or

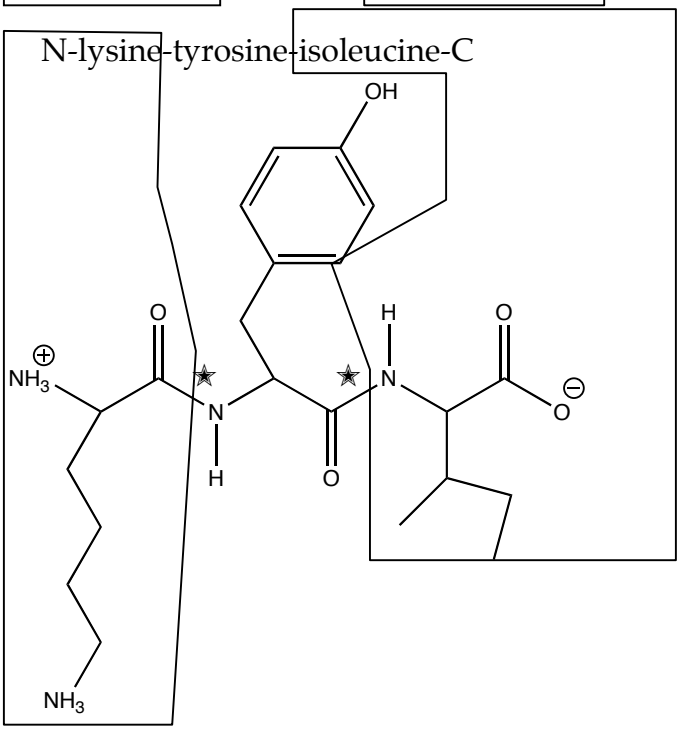
Both are leucine.

(C3) Computer-Aided Problems 3

a)

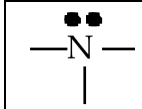
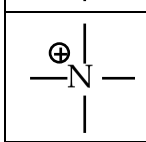
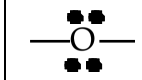
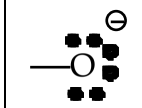
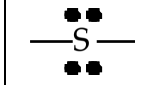


b)



(1.2) Noncovalent bonds and forces

(1.2.1)

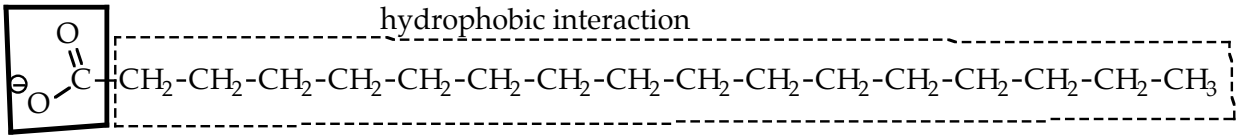
	Is the bond polar or nonpolar?	**Hydrophobic or hydrophilic?**	**Ionic bond?**	**Hydrogen bond?**	**Hydrophobic interactions?**
C–S	nonpolar - roughly equal electronegativities	nonpolar, so hydrophobic	not charged, so NO	no N or O, so NO	hydrophobic, so YES
P–O	polar - different electronegativities	polar, so hydrophilic	if O is charged, YES; otherwise, NO	if O: or O-H, YES; otherwise, NO	hydrophilic, so NO
S–H	nonpolar - similar electronegativities	nonpolar, so hydrophobic	not charged, so NO	no N or O, so NO	hydrophobic, so YES
S–O	polar - different eletronegativities	polar, so hydrophilic	if O is charged, YES; otherwise, NO	if O: or O-H, YES; otherwise, NO	hydrophilic, so NO
	an atom, so neither polar nor nonpolar	can make H bonds, so hydrophilic	not charged, so NO	lone pair on N (N:), so YES (acceptor)	hydrophilic, so NO
	an atom, so neither polar nor nonpolar	charged, so hydrophilic	N is charged, so YES	no N:, so NO (not to the N atom)	hydrophilic, so NO
	an atom, so neither polar nor nonpolar	can make H bonds, so hydrophilic	not charged, so NO	lone pair on O (O:), so YES	hydrophilic, so NO
	an atom, so neither polar nor nonpolar	charged, so hydrophilic	O is charged, so YES	lone pair on O (O:), so YES	hydrophilic, so NO
	an atom, so neither polar nor nonpolar	no charge or H bonds possible, so hydrophobic	not charged, so NO	no N or O, so NO	hydrophobic, so YES

(1.2.2)

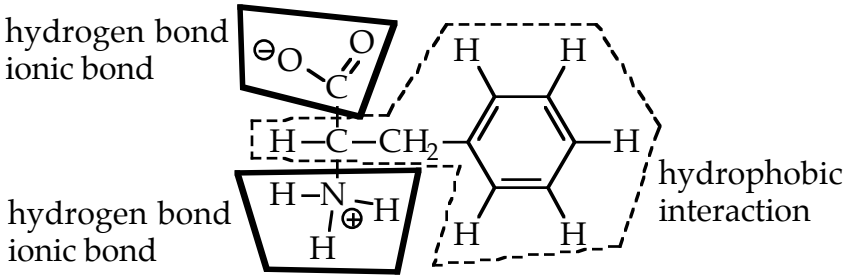
The toes of geckos adhered equally well to neutral, strongly hydrophobic, and strongly hydrophilic surfaces. If ionic or hydrogen bonds were the noncovalent forces involved, then you would not expect a gecko toe to stick to neutral or strongly hydrophobic surfaces. If hydrophobic interactions were the noncovalent forces, then you would not expect a gecko toe to stick to strongly hydrophilic surfaces. The adhesive properties of gecko toes are due to the van der Waals forces between the molecules on the millions of tiny hairs and the molecules of the surface. The millions of tiny hairs on the gecko toe increase the surface area and increase the number of van der Waals forces and increase the adhesive properties.

(1.2.3)

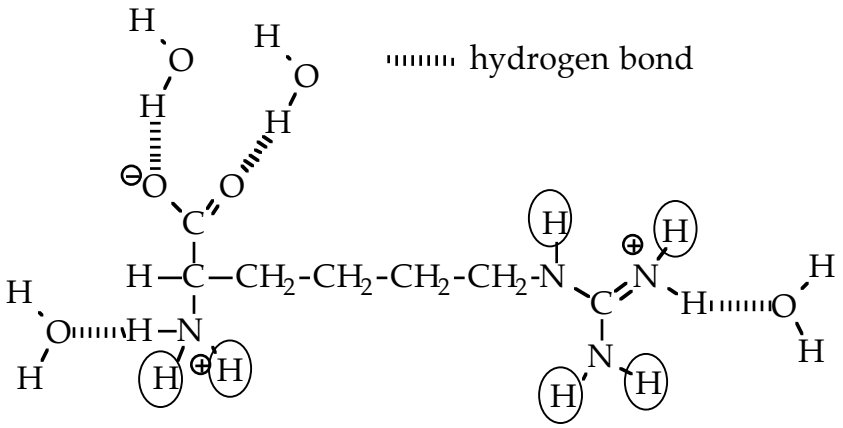
a) soap



b)



(1.2.4)



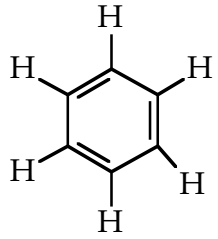
similar for all circled hydrogens

(1.2.5)

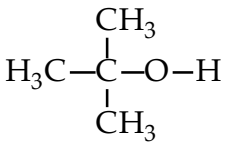
Part	Could this part form **ionic bonds** with another molecule?	Could this part form **hydrogen bonds** with another molecule?	Could this part form a **hydrophobic interaction** with another molecule?
(i)	YES. This is a nitrogen with a positive charge.	YES. It is a hydrogen donor, so it could form an H bond with an H acceptor.	NO. This is a polar hydrophilic part of the molecule
(ii)	NO. This is an oxygen with no charge, so it cannot make ionic bonds.	YES. It has two lone pairs of electrons (hydrogen acceptors), so it could make an H bond with an H donor.	NO. This is a polar hydrophilic part of the molecule
(iii)	NO. There is no charge to this part of the molecule.	NO. This part is exclusively carbon and hydrogen. It is therefore nonpolar and cannot participate in H bonding.	YES. This part is exclusively carbon and hydrogen. It is therefore nonpolar and a hydrophobic interaction is possible with a suitable partner.

(1.2.6)

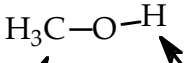
most hydrophobic



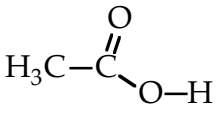
only nonpolar bonds, therefore very hydrophobic



has more hydrophobic parts than



most hydrophilic



has more hydrogen-bonding potential than

(C4) Computer-Aided Problems 4

1)

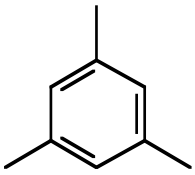
a) The simplest molecule is #2. Compared with #2, #1 has an added CH_2-OH. The added CH_2 is hydrophobic while the -OH is hydrophilic. Since polar groups like -OH contribute more to the hydrophilicity of a molecule on a per-atom basis than small hydrophobic groups, #1 is more hydrophilic than #2. Molecule #3 is more hydrophilic than #1 since it has an added charged group. Therefore, the most hydrophobic is #2, #1 is intermediate, and #3 is the most hydrophilic.

b) Using the Molecular Calculator, you should get:

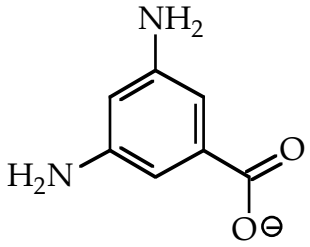
Molecule	**logP**
1	1.211
2	2.022
3	–1.798

Although answers for (c through e) will vary, a possible answer is given for each.

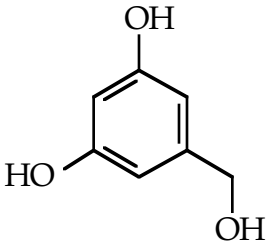
c) The following molecule is more hydrophobic than #2 because it has more hydrophobic CH_3 groups attached. Its logP is 2.7:



d) The following molecule is more hydrophilic than #3 because it has more hydrophilic -NH_2 groups attached. Its logP is –3.432:



e) The following molecule is more hydrophilic than #1 because it has more polar -OH groups; it is less hydrophilic than #3 because it is not charged. Its logP is 0.399:



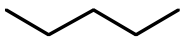
2)

a) This is one of the many possible answers for this question. In this case, R was a butyl group ($CH_3CH_2CH_2CH_2$-); this gives four structures:

HYDROPHOBIC

R-CH_3

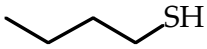
C_5H_{12}



logP = 2.13

R-SH

$C_4H_{10}S$

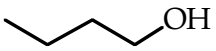


logP = 2.191

HYDROPHILIC

R-OH

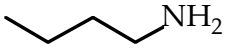
$C_4H_{10}O$



logP = 0.64

R-NH_2

$C_4H_{11}N$

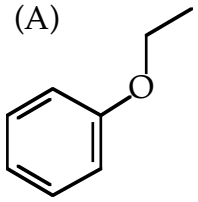


logP = 0.573

b) Since C-C, C-H, C-S, and S-H bonds are nonpolar, adding these to a molecule would make it more hydrophobic. Since O-H and N-H bonds are polar, adding these to a molecule would make it more hydrophilic.

3)
a) There are many possibilities; here is one pair of molecules:

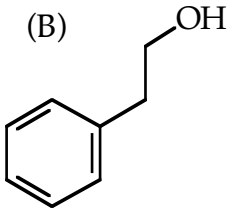
(A)



$C_8H_{10}O$

logP = 2.36

(B)



$C_8H_{10}O$

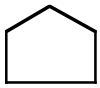
logP = 1.369

b) Both a C-O-C and a C-OH group have two free lone electron pairs on their oxygen atoms that are capable of making hydrogen bonds. However, the H atom of the -OH group can participate in hydrogen bonds that the C-O-C group cannot. Therefore, the alcohol form (B) would be expected to be more hydrophilic than the ether form (A).

c) The logP values agree with the prediction.

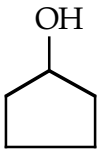
4)
a) through c) There are many possible answers; here is one set of molecules:

Original Molecule



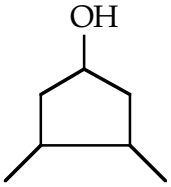
logP = 1.79

Add an -OH



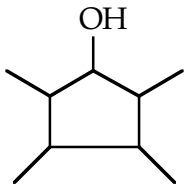
logP = 0.76

Add two -CH_3's



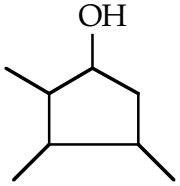
logP = 1.35
not quite enough

Add four -CH_3's



logP = 1.948
a little too much

Add three -CH_3's

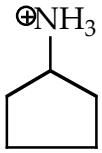
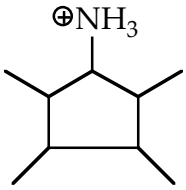
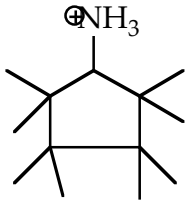
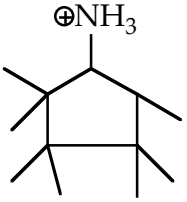


logP = 1.651
pretty close

It took between three and four -CH_3 groups to overcome the effect of the -OH. You should get similar results. This shows that, per atom, hydrophilic groups contribute more than hydrophobic groups.

5)

a) through c) There are many possible answers; here is one set of molecules:

Original Molecule	Add an $-NH_3^+$	Add four $-CH_3$'s	Add eight $-CH_3$'s	Add seven $-CH_3$'s
				
logP = 1.79	logP = -0.0773	logP = 0.415 not quite enough	logP = 1.995 a little too much	logP = 1.6 pretty close

It took about seven $-CH_3$ groups to overcome the effect of the $-NH_3^+$. You should get similar results. This shows that charged groups are more hydrophilic than uncharged polar groups.

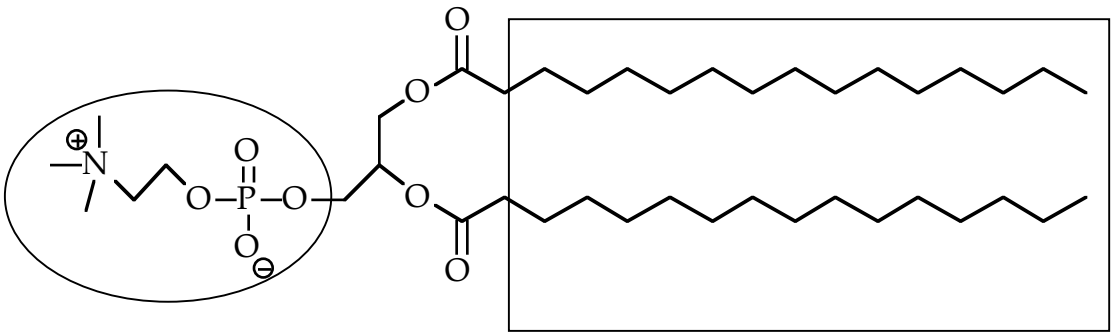
(2) MACROMOLECULES

(2.1) Lipids and phospholipids

(2.1.1)
A saturated fatty acid can form a solid at room temperature. The hydrocarbon chain of a saturated fatty acid contains no C=C bonds, so the molecules can align and pack together in an orderly fashion. The C=C bonds in the hydrocarbon chain of an unsaturated fatty acid prevent this orderly arrangement and result in a lower melting point.

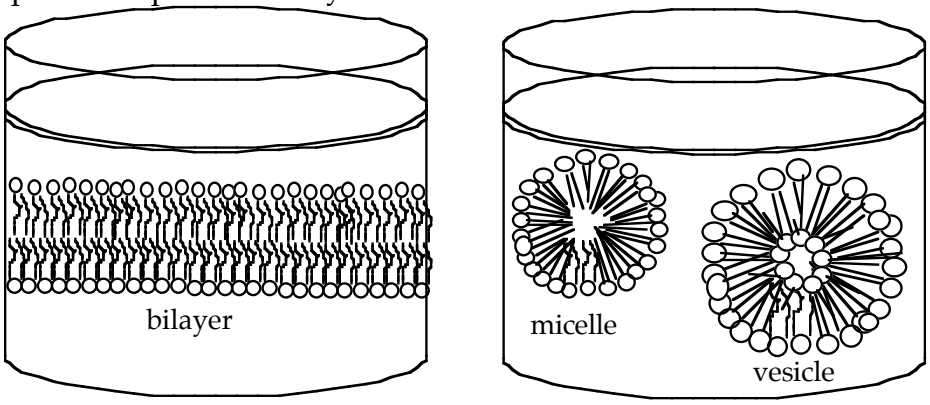
(2.1.2)
Phospholipids are a major component of <u>membranes</u>.

A phospholipid contains both polar and nonpolar domains. Circle the polar domain. Box the nonpolar domain.



The environment within the cell is aqueous and therefore hydrophilic. The arrangement of phospholipids shown had the hydrocarbon tails exposed to this environment. These hydrocarbon tails are extremely hydrophobic.

(2.1.3)
Phospholipids can spontaneously form the three different structures below.



(2.2) Nucleic acids

(2.2.1)
The first molecule is deoxyribonucleic acid (DNA). DNA can be distinguished from ribonucleic acid (RNA) by the absence of an OH group on the 2′ carbon. DNA is the genetic material.

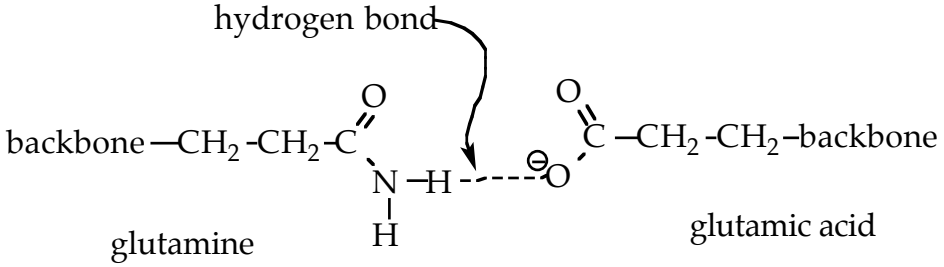
The second molecule is RNA. There are several different types of RNA in the cell.
Messenger RNA (mRNA) is an RNA copy of a gene that is used in protein synthesis.
Ribosomal RNA (rRNA) forms part of the structure of the ribosomes.
Transfer RNA (tRNA) acts as an adaptor molecule. Each tRNA recognizes a codon in the mRNA and carries the appropriate amino acid to the ribosome.

(2.3) Polypeptides and proteins, background

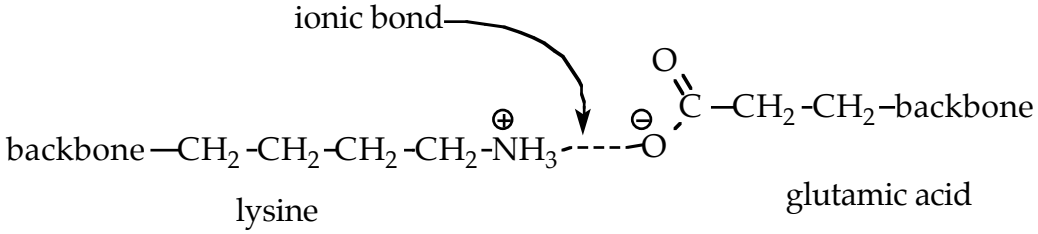
(2.3.1)
Remember that the problem asked only about the side chains, so don't consider the $-NH_3^+$ or $-COO^-$ groups.

a) Glutamine's side chain has both H bond donors (2 N-H's) and H bond acceptors (=O: and N:) so it can form H bonds with any amino acid that has either a hydrogen donor or a hydrogen acceptor.

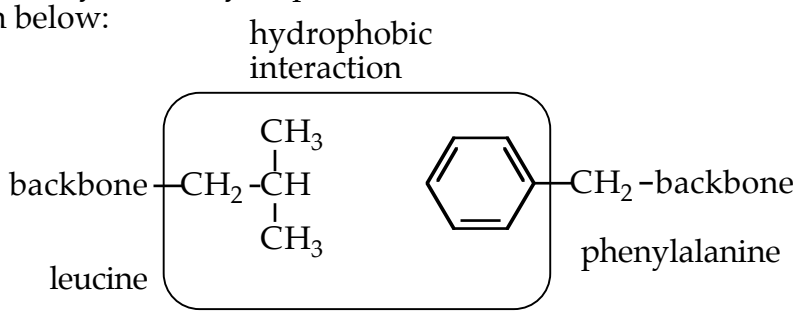
Example: Glutamic acid has only hydrogen bond acceptors on its side chain. It could form many hydrogen bonds with the side chain of glutamine; here is one possibility:



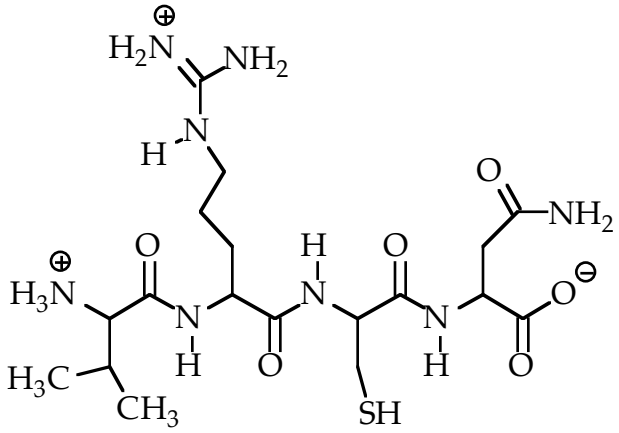
b) Lysine has a (+) charge, so it can only form an ionic bond with a (–) charge. The only amino acids with (–) charged side chains are glutamic acid and aspartic acid. Glutamic acid can form only one ionic bond with the side chain of lysine. This is shown below:



c) Leucine has a hydrophobic side chain. It can form a hydrophobic interaction with any hydrophobic part of a side chain. Phenylalanine has a completely hydrophobic side chain and can easily form a hydrophobic interaction with the side chain of leucine. This is shown below:



(2.3.2)
N-Val-Arg-Cys-Asn-C



(C5) Computer-Aided Problems 5

1) a) The complete secondary structure is:

1 – 2: random coil
3 – 11: alpha helix
12 – 13: random coil
14 – 20: beta sheet
21 – 23: turn
24 – 27: beta sheet
28 – 30: turn
31 – 34: beta sheet
35 – 38: random coil
39 – 50: alpha helix
51 – 53: random coil
54 – 55: turn
56 – 58: beta sheet
59: random coil
60 – 80: alpha helix
81: random coil

82 – 90: alpha helix
91 – 92: random coil
93 – 106: alpha helix
107: random coil
108 – 113: alpha helix
114: random coil
115 – 123: alpha helix
124 –125: random coil
126 – 134: alpha helix
135 – 136: random coil
137 – 141: alpha helix
142: random coil (it can sometimes be hard to see this amino acid)
143 – 155: alpha helix
156 – 158: random coil
159 – 162: turn
163: random coil

b) Depending on how you count them, there are approximately three amino acids per turn of the alpha helix. The actual value, as measured from many proteins, is 3.6.

c) The four sections interact as two pairs:
(15-17):(56-58) are antiparallel
(24-27):(31-34) are also antiparallel

(2.4) Polypeptides and proteins, interactions

(2.4.1)

a)

Gln_{57}	hydrophilic and polar		Pro_{48}	hydrophobic
Leu_{60}	hydrophobic		Ile_{46}	hydrophobic
Lys_{67}	hydrophilic and charged		Phe_{83}	hydrophobic
Glu_{71}	hydrophilic and charged		Arg_{34}	hydrophilic and charged

b)

MHC II side chains	Interaction with opposite side chain of TSST
Gln_{57}	van der Waals
Leu_{60}	van der Waals + hydrophobic
Lys_{67}	van der Waals
Glu_{71}	van der Waals + ionic + hydrogen

c) There are a number of different answers for this part. The idea is to make another strong bond. The strongest bond to add would be an ionic bond. For example, if you change Phe_{83} of TSST to an aspartic acid, the Asp_{83} of TSST would now be able to form an ionic bond with Lys_{61} of MHC II.

d) Changing Arg_{34} to Gln prevents the formation of an ionic bond with Glu_{71} of MHC II. The interaction between Gln_{34} of TSST and Glu_{71} of MHC II would be a hydrogen bond. It may be that an H bond at that position is not strong enough to hold the proteins together.

e) A glutamic acid at position 34 instead of an arginine replaces a positively charged amino acid with a negatively charged one. The glutamine at position 71 of MHC II is also negatively charged. The two (–) charges will now repel each other. The remaining van der Waals interactions must not be strong enough to overcome the repulsion.

f) van der Waals interactions occur between any two nonbonded atoms. Replacing an isoleucine with leucine is unlikely to change the strength of the MHC II-TSST interaction, so they should still bind.

g) With these substitutions, one ionic bond and a van der Waals force have been replaced with two H bonds. It must be that the two H bonds are strong enough to hold the two proteins together.

h) One possibility would be to change Glu_{71} to lysine or arginine, thus restoring the ionic bond.

(2.4.2)

a) The interaction between the side chain of valine #6 and the side chains of phenylalanine #85 and leucine #87 is a hydrophobic interaction.

b) The aggregation is driven by an interaction between the side chain of amino acid #6 (valine) of one hemoglobin molecule with a pocket formed by the side chains of amino acids #85 (phenylalanine) and #87 (leucine) of another hemoglobin molecule. The amino acid #6 in Hb^{+} (wild-type) is glutamic acid. Glutamic acid is a charged amino acid and cannot participate in a hydrophobic interaction with amino acids #85 and #87 of another hemoglobin molecule.

c) Phenylalanine is hydrophobic like valine, so it should be able to interact with the side chains of amino acids #85 (phenylalanine) and #87 (leucine) of another hemoglobin molecule. However, phenylalanine is larger than valine.

i) If Hb^{Phe} **<u>does not</u>** form polymers under any circumstances, it is possible that the larger phenylalanine cannot fit into the pocket formed by the side chains of amino acids #85 (phenylalanine) and #87 (leucine) of another hemoglobin molecule.

ii) If Hb^{Phe} **<u>does</u>** form polymers with the same general structure as polymers of Hb^{S}, then it must fit into the pocket formed by the side chains of amino acids #85 (phenylalanine) and #87 (leucine) of another hemoglobin molecule and interacts in the same manner as valine.

(2.4.3)
a) There are two models for why these amino acid changes result in an inactive enzyme: 1) the new amino acid side chain is charged instead of neutral; or 2) the side chain of the new amino acid is larger than that of glycine.

The first data set is ambiguous, because the side chains of the added amino acids are both big and charged. However, the second data set shows that substituting valine, whose side chain is larger than glycine but uncharged, results in an inactive enzyme. This indicates that it is the size of the amino acid side chain at position A that matters, not the charge (model ii). If this is the case, it is possible that the binding pocket can accommodate only an amino acid with a small side chain like glycine. Substituting an amino acid with a larger side chain alters the pocket and changes the conformation of the molecule so that it is no longer able to function.

b) A possible explanation for mutant 6 is that the amino acids at positions A and B interact somehow, so that changes in one can compensate for changes in the other. For example, perhaps when the enzyme is in its active form, side chains of both A and B lie near each other in the pocket within the enzyme. Mutant 1 (Gly ⇒ Glu) is inactive because the Glu side chain is too big to fit in the pocket where Gly normally sits. The effects of this can be partially reversed by changing Tyr ⇒ Cys, which opens up more space in the pocket to accommodate the side chain of Glu.

c) In this case, it appears that reducing the size of a side chain in the pocket (Tyr ⇒ Cys) also results in formation of an inactive enzyme. Therefore, the pocket has to be a certain size for the enzyme to function: Gly ⇒ Val overfills it and Tyr ⇒ Cys underfills it. Given the compact nature of protein structures, these results are not surprising.

(2.4.4)
a) Each subunit is a polypeptide, so eight polypeptides compose the nucleosome complex.

b) Quarternary structure is the association of different polypeptides or subunits. The nucleosome has eight interacting polypeptides so it has quarternary structure.

c) The amino acids in the sequence are hydrophobic. You would expect to find this stretch of amino acids to be buried in the hydrophobic interior of the complex.

d) The regions high in the positively charged amino acids lysine and arginine would likely be on the surface of the complex. Lysine and arginine can form ionic bonds with the negatively charged backbone of the DNA.

e)

Treatment	Effect on nucleosome complexes list appropriate number(s) from above
Disrupt hydrogen bonds	4, 2
Disrupt ionic bonds	2
Disrupt peptide bonds	2, 4, 5

f)

Treatment	Effect on structure of DNA double helices list appropriate number(s) from above
Disrupt hydrogen bonds	2
Disrupt ionic bonds	1
Disrupt peptide bonds	1

(2.4.5)

a)

	Amino acids	Strongest interaction
i)	Ser and Gln	Hydrogen bond
ii)	Ala and Met	van der Waals forces
iii)	Phe and Tyr	van der Waals forces

b) The α-helical structure of each polypeptide is due to a large number of hydrogen bonds. The interaction between the two polypeptides is based on van der Waals forces and a hydrogen bond. Low heat is sufficient to disrupt weak binding between the two α-helices, yet secondary structure is maintained. High heat will disrupt all hydrogen bonds and produce elongated peptides.

c) Under some conditions, changing both of these residues to cysteine will increase the stability of the interaction. Two closely opposed cysteine residues are capable of forming a type of covalent bond called a disulfide bond.

(C6) Computer-Aided Problems 6

1) Hydrophobic/hydrophilic

a) The exterior is mostly hydrophilic. Surprisingly, there are many hydrophobic side chains on the surface. This is likely because, on a per-atom basis, hydrophilic elements make a much greater contribution to the molecule than the hydrophobic parts. To put it another way, the hydrophobic effect is so weak that only a few hydrophilic groups are required to make the protein soluble and stable.

b) The interior of the protein is almost exclusively hydrophobic. This is not surprising as even a little hydrophilic character would prevent a side chain from remaining in the interior.

c) There are nine valines in lysozyme:

Valine	Position of side chain
57	pokes out into water
71	small exposure
75	pokes out into water
87	small exposure
94	pokes out into water
103	very small exposure
111	totally buried
131	pokes out into water
149	totally buried

Thus, only two of nine are totally in the interior of the protein. This is consistent with part (a). The surface of many proteins is a mosaic of hydrophobic and polar amino acids.

d) None of the 13 lysines are completely in the interior of the protein; all of the $-NH_3^+$ groups are exposed to the water. This is exactly as expected.

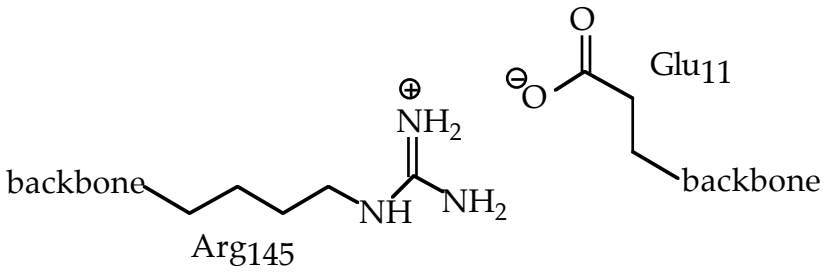
2) Side-chain interactions

a) Glu_{11} and Arg_{145}.

i) Glu has a (–) charge and H-bond acceptors (lone pairs on oxygen atoms); Arg has a (+) charge and H-bond acceptors (lone pairs on nitrogen atoms) as well as H-bond donors (hydrogen atoms covalently bonded to nitrogen atoms). Therefore, ionic bonds, hydrogen bonds, and van der Waals interactions are possible. Since both are highly hydrophilic, a hydrophobic interaction is not possible.

ii) Ionic bond is the strongest.

iii) The structure of the side chains looks like this:



Note that the (+) charge could be on any of the N's in arginine and the (–) charge could be on either of the oxygens in glutamic acid. There is not enough information in the structure to tell which is correct. [In fact, the (+) charge is distributed over the three N's and the C at the end of the arginine and the (–) charge is distributed over the two O's and the C in glutamic acid. This is called resonance, but it is beyond the scope of this book.]

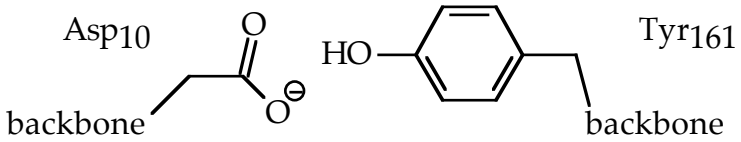
iv) Since the (+) and (–) charges are in close proximity, an ionic bond is the strongest possible interaction between these side chains.

b) Asp_{10} and Tyr_{161}

i) Asp has a (–) charge and H-bond acceptors (lone pairs on oxygen); it has only one CH_2 in its side chain, not enough for a hydrophobic interaction. Tyr has an –OH, which can serve as an H-bond donor and an H-bond acceptor; it also has a large hydrophobic ring. Since only one is charged, there cannot be an ionic bond. Since Tyr has an H-bond donor and Asp has H-bond acceptors, an H bond is possible. Since Asp is not hydrophobic, a hydrophobic interaction is not possible. van der Waals is always possible.

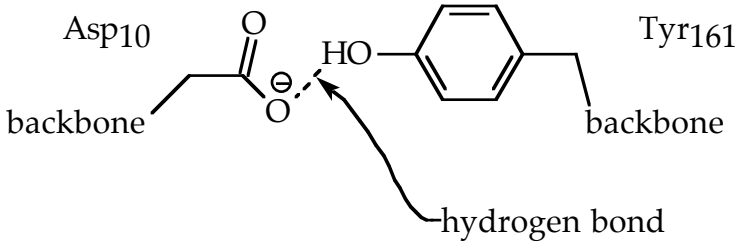
ii) H bond is the strongest possible.

iii) In the protein, the side chains are oriented like this:



Note that the (–) charge could be on either of the oxygens in aspartic acid. For our purposes, it does not matter.

iv) Since the H-bond donor (the H on the OH in Tyr) points toward the H-bond acceptor (the lone pair on the O in Asp), a hydrogen bond is the most likely bond. It would look like this:

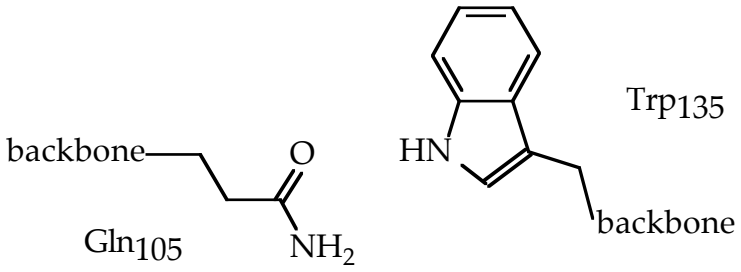


c) Gln_{105} and Trp_{138}

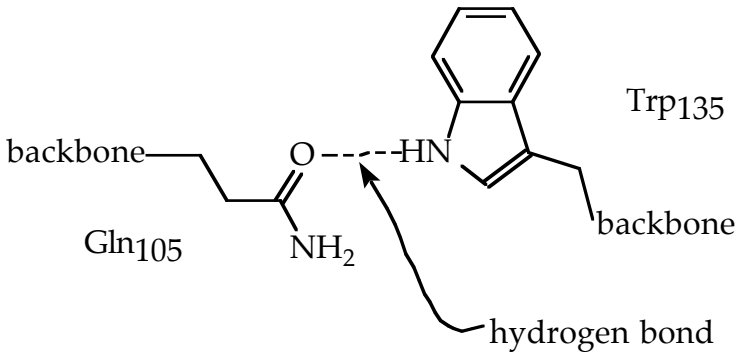
i) Glutamine's side chain has no charge. It has both H-bond donors (H's on N) and H-bond acceptors (lone pairs on N and O). It has a short chain of CH_2's, so a hydrophobic interaction is possible, but unlikely. Tryptophan's side chain has no charge. It has both an H-bond donor (H on N) and an H-bond acceptor (lone pair on N). It also has a highly hydrophobic ring system. Therefore, an ionic bond is impossible; a hydrogen bond is possible; hydrophobic interaction is possible; and van der Waals is possible.

ii) The strongest is H bond; many such bonds are possible.

iii) In the protein, the side chains of the two amino acids are oriented like this:



iv) Since the H-bond donor (the H on the N in Trp) points right at an H-bond acceptor (the lone pair on the O in Gln), an H bond is very likely. It would look like this:

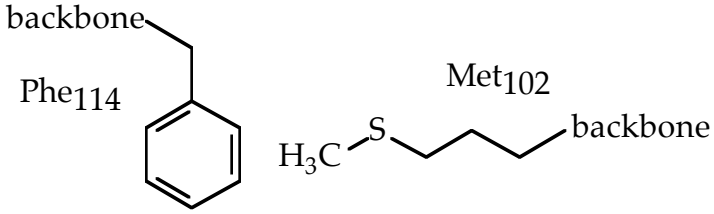


d) Met_{102} and Phe_{114}

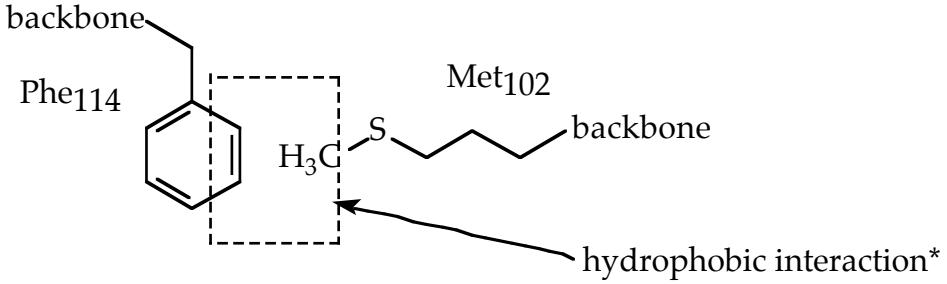
i) Met has no charge or H-bond possibilities; it has a large hydrophobic side chain. Phe has similar properties. Therefore, only a hydrophobic interaction or van der Waals is possible.

ii) The strongest is hydrophobic interaction.

iii) In the protein, the side chains are oriented like this:



iv) Based on the structure, the strongest possible interaction is a hydrophobic interaction. It could be shown like this:



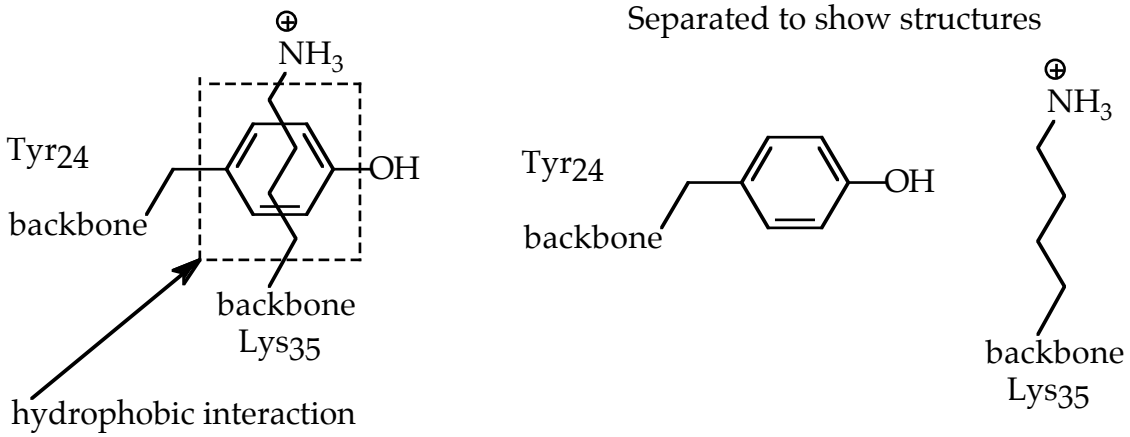
*It is shown as a dashed box rather than a bond because it is not a true bond between a pair of atoms. It is an interaction between these groups and the water surrounding the protein.

e) Tyr_{24} and Lys_{35}

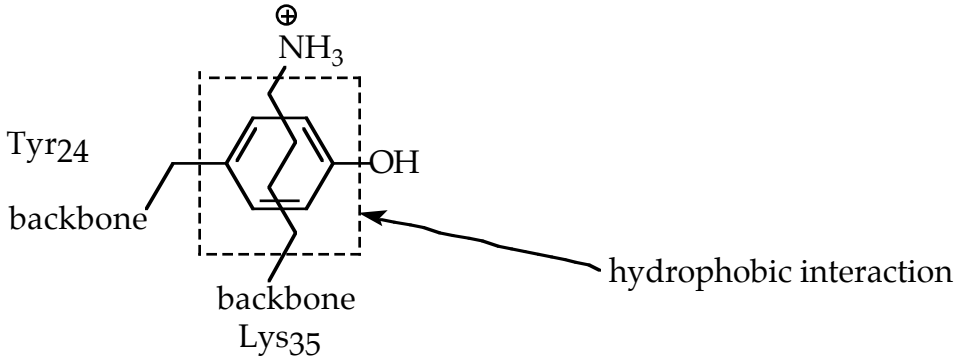
i) Tyrosine's side chain has no charge. It does have an H-bond donor and an H-bond acceptor. It also has a large hydrophobic ring. Lysine's side chain has a (+) charge. It also has H-bond donors (the H's on the N^+) but no H-bond acceptor (no lone pair on the N). It also has a somewhat hydrophobic set of four CH_2's. Therefore, ionic bonds are not possible. A hydrogen bond, hydrophobic interaction, or van der Waals is possible.

ii) The strongest is an H bond.

iii) In the protein, the side chains are oriented like this:



iv) Since the H-bond donor (H's on the N^+ of lys) is nowhere near the H-bond acceptor (lone pair on O of Tyr), an H-bond is not possible. However, the hydrophobic parts of the two molecules are very close together. Therefore, a hydrophobic interaction is the most likely interaction in this case. This is why it is important to look at the relative orientation of the side chains, rather than just that they are neighbors. The interaction could be shown like this:



3) Effects of mutations on protein structure

a) Glu_{11} normally makes an ionic bond with Arg_{145}. If Glu_{11} were replaced by Arg, that Arg's (+) charge would be right next to the (+) charge of Arg_{145}. Given that like charges repel, this would destabilize the protein, making it not fully active.

b) Glu_{11} normally makes an ionic bond with Arg_{145}. If Glu_{11} were replaced by Phe, which cannot make ionic bonds because it is not charged, this ionic bond would not be present in the mutant protein. Without this interaction to hold it in the proper shape, the mutant protein is not fully active.

c) Glu_{11} normally makes an ionic bond with Arg_{145}. If Glu_{11} were replaced by Asp, you might expect that this ionic bond would still be possible. However, Asp's side chain is not as long as that of Glu. Looking at the protein's structure, Glu_{11}'s side chain is stretched out very straight as it reaches out to Arg_{145}. If the side chain were shorter, the

(–) charge on Asp would not be close enough to make an ionic bond to Arg_{145}. This weakens the structure enough so that the protein is not fully active.

d) Arg_{145} normally makes an ionic bond with Glu_{11}. If Arg_{145} were replaced by Ser, an ionic bond is no longer possible. Although Ser could possibly form an H bond [although it may be too short – see part (c)], this is weaker than an ionic bond. This weaker bond, if it exists, must not be strong enough to maintain the proper structure of the protein and it is not fully active.

e) Arg_{145} normally makes an ionic bond with Glu_{11}. Since both His and Lys have (+)-charged side chains, you would expect that either substitution would have no effect on the protein's activity. The only relevant difference between His and Lys is that Lys's side chain is substantially longer than that of His. It is likely, then, that although Lys still has a (+) charge, it is too long to make a proper ionic bond with Glu_{11} and so the resulting protein is not fully active. His, on the other hand, is the right size, so the resulting protein is fully active.

f) Tyr_{161} normally forms an H bond with Asp_{10}. If Tyr_{161}were replaced by Ser, an H bond would be possible, based on the structures alone. However, if you look at the structures of the side chains in the protein, the side chain of Tyr_{161} is stretched far out from the backbone. Since the side chain of Ser is much shorter than that of Tyr, the Ser cannot reach to Asp_{10} to make an H bond and, as a result, the protein lacks this crucial H bond and is not fully active.

g) Asp_{10} normally forms an H bond with Tyr_{161}. If Asp_{10} were replaced by Glu, an H bond will still be possible. Although the side chain of Glu is one carbon longer than that of Asp, this does not affect the H bond, and the altered protein is still fully active.

h) Gln_{105} normally makes an H bond with Trp_{138}. If Gln_{105} were replaced by Glu, the H bond could still be formed and the length of the new side chain would be exactly the same as the normal protein. This explains why the altered protein is fully active.

i) Gln_{105} normally makes an H bond with Trp_{138}. If Gln_{105} were replaced by Leu, that H bond would be impossible. Therefore, you'd expect that the mutant protein would not be active. Perhaps, the Leu makes a hydrophobic interaction with Trp_{138} and this is enough to stabilize the protein.

j) Met_{102} normally makes a hydrophobic interaction with Phe_{114}. If Met_{102} were replaced by Glu, Lys, or Arg, this type of interaction would not be possible. In addition, based on the view shown in the CD-ROM for this problem, the side chain of Met_{102} is in the hydrophobic core of the protein. Therefore, a hydrophilic amino acid at this position would be expected to seriously disrupt the protein's structure as it would "prefer" to be on the surface of the protein. The absence of the hydrophobic interaction, or the disruption described, is enough to cause the protein to not be fully functional.

k) Lys_{35} normally makes a hydrophobic interaction with Tyr_{24}. Based on the view shown in the CD-ROM for this problem, the side chain of Lys_{35} is on the surface of the protein. Because Lys_{35} can be replaced by any other amino acid, we must conclude that this hydrophobic interaction is not important for the protein's structure and function.

Furthermore, substitutions to hydrophilic amino acids are tolerated because their side chains will be on the surface of the protein; substitutions to hydrophobic amino acids are also tolerated because a small number of surface hydrophobics are always tolerated.

l) Phe_{67} is in the middle of an alpha-helical section of the protein. Although Pro is hydrophobic like Phe, Pro has a highly constrained backbone that tends to destabilize α-helices. Presumably, changing Phe_{67} to Pro disrupts the α-helix and renders the protein inactive.

(2.5) Polypeptides and proteins, binding sites

(2.5.1)

a) The closest part of Molecule X to Glu_{75} is the $-NH_3^+$. It has a (+) charge and H-bond donors; it is therefore hydrophilic, so it cannot participate in a hydrophobic interaction. The side chain of Glu_{75} is uncharged so an ionic bond between Molecule X and Glu_{75} is not possible. The side chain of Glu_{75} has both H-bond donors and acceptors; the possible interactions are H bonds and van der Waals forces. The strongest of these is the H bond.

b) The closest parts of Molecule X to Ile_{147} are the CH's of the ring. They are hydrophobic and uncharged and cannot make ionic or H bonds. Ile_{147} is uncharged as well. The possible interactions are, therefore, hydrophobic interaction or van der Waals. The strongest is hydrophobic interaction.

c) The closest part of Molecule X to Lys_{302} is the $-COO^-$ group. It is highly charged and cannot participate in hydrophobic interactions. The closest part of Lys_{302} to Molecule X is the $-NH_3^+$. It has a (+) charge and H-bond donors. The possible interactions are, therefore, ionic bond, H bond, and van der Waals. The strongest is ionic bond.

d) Glutamine and asparagine have essentially identical bonding capabilities so that you would expect this change to have no effect on the bonding of Molecule X to the protein. The only difference between the two side chains is that asparagine's side chain is one carbon shorter than glutamine's. Presumably, this shorter side chain cannot reach to Molecule X and form the necessary H bond. As a result, Molecule X no longer binds to the protein.

e) Glutamic acid has a negatively charged side chain; lysine has a positively charged side chain. In the altered protein, there will be two negative charges near each other. Since like charges repel, you would not expect Molecule X to bind the altered protein.

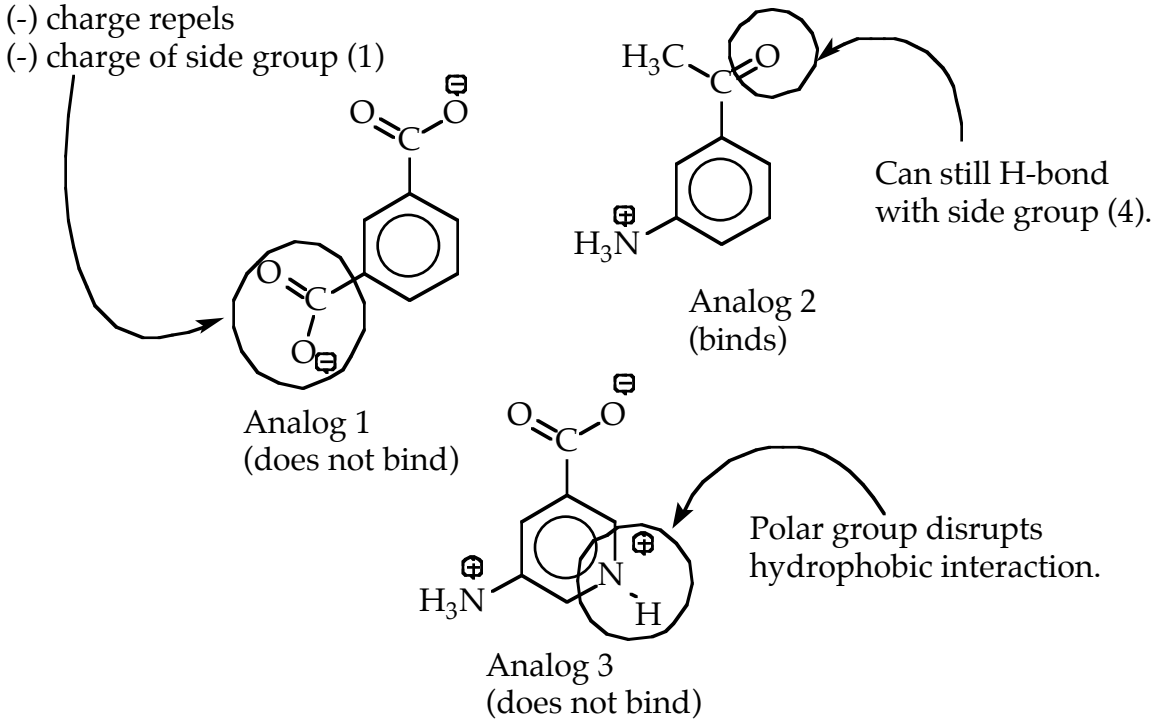
(2.5.2)

a) (1) aspartic acid (2) cysteine (3) valine (4) asparagine

b)

Group	Interaction(s) of Group with Substrate	Classification of Group
(1)	ionic (hydrogen also possible)	hydrophilic, charged
(2)	all nonpolar bonds so hydrophobic interactions or van der Waals forces	hydrophobic
(3)	all nonpolar bonds so hydrophobic interactions or van der Waals forces	hydrophobic
(4)	hydrogen bond	hydrophilic, polar

c)



d) Groups (1), (2), and (3) make the strongest interactions possible given the structure of the substrate. That is:

• Group (1), whatever it is, will be able to interact with the $-NH_4^+$ group of the substrate–this can make H bonds or ionic bonds–the strongest are ionic. The normal enzyme makes an ionic bond, so this cannot be stronger.

• Group (2), whatever it is, will be able to interact with some C-C and C-H parts of the substrate–a very nonpolar region. Therefore, only hydrophobic interactions are possible with that part of the substrate and that is what is already present.

• Group (3), whatever it is, will be able to interact with some C-C and C-H parts of the substrate–a very nonpolar region. Therefore, only hydrophobic interactions are possible with that part of the substrate and that is what is already present.

• Group (4), whatever it is, will be able to interact with the $-COO^-$ region of the substrate, which can make a hydrogen bond or an ionic bond. In the normal protein, it

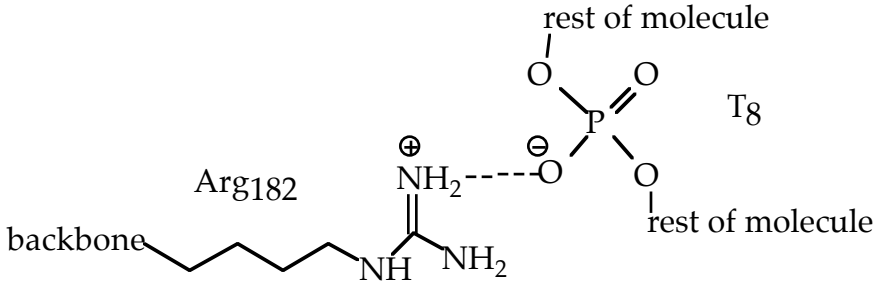
is a hydrogen bond; an ionic bond would be stronger. To make an ionic bond, you'd need a positively charged amino acid at position (4). If the amino acid at position (4) was changed to lysine, arginine, or histidine, the binding could be strengthened.

(C7) Computer-Aided Problems 7
1) Molecules in three dimensions

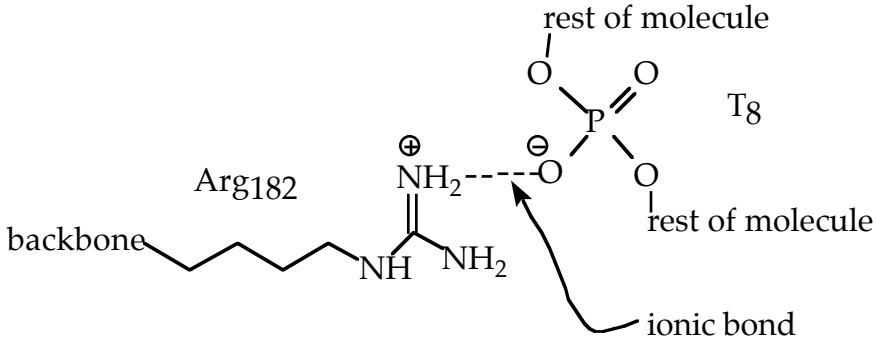
a) There is no answer for this part.

b)

i) Arg_{182} and T_8. The structures of the relevant parts are:

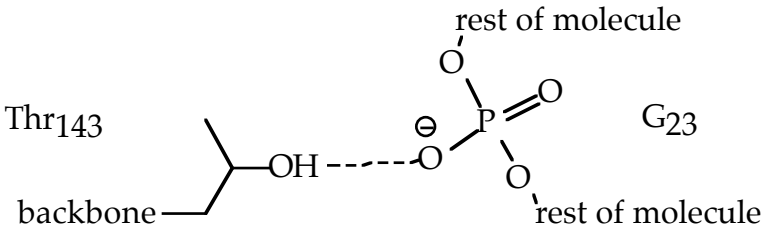


ii) Note that the structure does not make it clear which N has the (+) charge in Arg, nor does it make clear which oxygen in the DNA is (–) charged. However, the strongest bond is achieved if the charges are as shown; thus, this is the most likely configuration. The closest part of the Arg has a (+) charge as well as H-bond donors and acceptors. The closest part of the DNA has a (–) charge and H-bond acceptors. Therefore, ionic bonds, H bonds, and van der Waals interactions are possible. An ionic bond is the strongest of these. It would be shown like this:

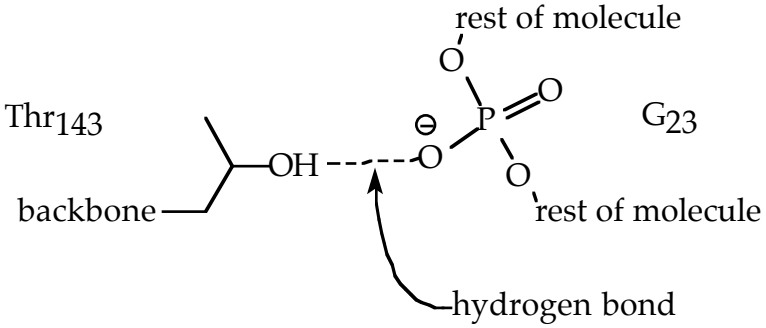


c)

i) Thr_{143} and G_{23}. The structures of the relevant parts are:

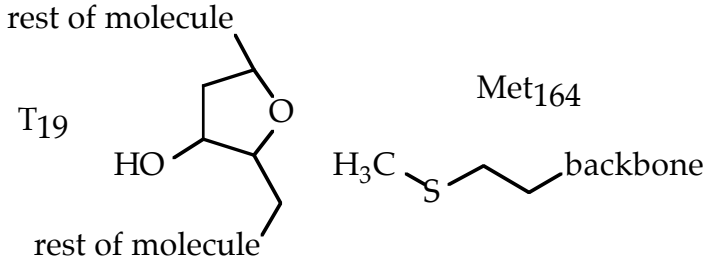


ii) Note that the structure does not make it clear which oxygen atom has the (–) charge. For the purposes of this problem, it does not matter, since either form has a lone pair H-bond acceptor. The closest part of the Thr is the –OH, which is an H-bond donor. Therefore, H bonds and van der Waals interactions are possible. The strongest is an H bond.



d)

i) Met_{164} and T_{19}. The structures of the relevant parts are:

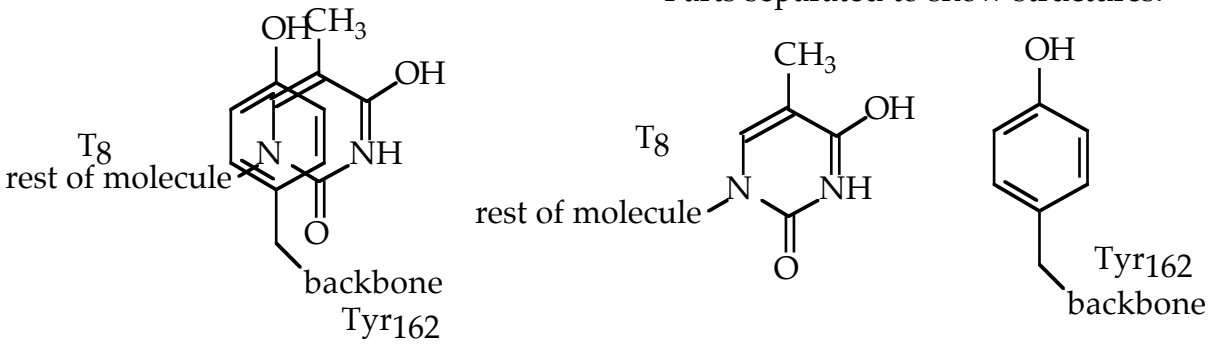


ii) The closest part of T_{19} has H-bond acceptors and some small hydrophobic parts. The closest part of Met_{164} is also hydrophobic. Therefore, a hydrophobic interaction or van der Waals forces are possible. The strongest is a hydrophobic interaction.

e)

i) Tyr_{162} and T_8. The structures of the relevant parts are:

Parts separated to show structures:



ii) Tyr has a large hydrophobic ring as well as H-bond donors and acceptors. Thymine has a hydrophobic ring as well as H-bond donors and acceptors. Looking carefully at the structures, the two –OH groups are not close enough to

form an H bond [see how close they are in part (c), especially using the spacefill view]. Therefore, the only possible bonds are hydrophobic interaction and van der Waals. Hydrophobic interaction is the stronger of the two.

f) Thr_{143} forms an H bond with G_{23}. Gln also has H-bond donors that could form an H bond with this part of G_{23}, resulting in a fully functional protein.

g) Met_{164} forms a hydrophobic interaction with T_{19}. Both Ile and Phe are also hydrophobic, so they could form a hydrophobic interaction with T_{19} as well, resulting in a fully functional protein.

h) Tyr_{162} forms a hydrophobic interaction with T_8. You would the expect that other hydrophobic amino acids (especially he) would be able to form a similar, if not identical, interaction. It must be that Tyr_{162} makes other essential interactions that require this amino acid to be Tyr.

(C8) Computer-Aided Problems 8

a) There is no answer for this part.

b) Going clockwise starting at His_{90}, the interactions are:

- His_{90}: H bond between H on N^+ of His (H-bond donor) and lone pair on O of drug (H-bond acceptor).
- Tyr_{355}: hydrophobic interaction or van der Waals between ring of Tyr and ring 1 of drug.
- Arg_{120}: H bond between H on N (or N^+) of Arg (H-bond donor) and lone pair on N (H-bond acceptor) in ring 2 of drug.
- Leu_{531}: hydrophobic interaction or van der Waals between $-CH_3$'s of Leu and CH of ring 2 of drug.
- Ser_{530}: hydrophobic interaction or van der Waals between CH_2 of Ser and ring 3 of drug.
- Phe_{381} and Tyr_{385}: hydrophobic interaction or van der Waals between rings of Phe and Tyr and ring 3 and the Fluorine atom of the drug.

Note that many of these interactions are van der Waals. Although they do not provide much bond strength, they do provide shape specificity (as you will see in the next few parts).

c)

i) No answer is required for this part.

ii) Since this is even larger than the $-CH_2CH_3$ group (which does not bind), you would not expect this to bind to COX-2. This is what the researchers observed.

d)

i) Part B of the molecule is right up against the side chain of Tyr_{355}. The data show that anything larger than a hydrogen atom will not fit in this space.

ii) $-CH_3$ should be larger than –Cl and smaller than $-CH_2OH$. Since both –Cl and $-CH_2OH$ are too large for the pocket, you would not expect the B = $-CH_3$ version of the drug to bind to COX-2. This is what the researchers found.

e)

i) The SO_2 group in part C of the drug does interact with His_{90}. The remainder of part C does not interact with any of the side chains shown. Changing $-CH_3$ to $-NH_2$ does not change the size of this part of the drug appreciably, nor does it interfere with the H bond between the oxygens in the drug and the side chain of His_{90}. Therefore, it is not surprising that this molecule binds.

ii) This version lacks the ability to form an H bond with the side chain of His_{90}. You would not expect it to bind COX-2; this is what the researchers found.

f)

i) Part D of the drug interacts with the side chains of Tyr_{385} and Phe_{381}. Based on these results, there is a medium-sized space for group D's atoms. Even very small group D's are tolerated, indicating that this van der Waals interaction is not critical.

ii) Since this group is going to be about the same size as $-SCH_2CH_3$ and the hydrophilic oxygen atom is tolerated in other derivatives, you would expect this derivative to bind COX-2 (it does).

g)

i) This is tricky, but (very roughly speaking), Celebrex would be closest to the molecule:

A = –H
B = –H
C = $-SO_2NH_2$
D = $-CH_3$

ii) You would not expect this to bind because derivatives where A = –H do not bind to COX-2. Perhaps the $-CF_3$ group makes a compensating van der Waals bond with another amino acid.

(3) ENERGY, ENZYMES, AND PATHWAYS

(3.1) Energy and enzymes

(3.1.1)

a) When calculating ΔG for reactions, you need consider only the energy states of the reactants and products. Since the ΔG difference between A and B is most negative, A $\Rightarrow$ B is the more spontaneous of the two reactions.

b) Though reaction 1 is more thermodynamically favorable, it has a huge energy barrier (a very positive ΔG) to overcome before changing to I_1. Having one step in a reaction that is very unfavorable will not change the energy of the overall reaction, but it can slow it down drastically. Reaction 2 has smaller energy barriers to overcome as it converts to I_3 and I_4, so even though its ΔG is less negative, it is likely to be faster than reaction 1.

c) The overall ΔG of the reaction would stay the same, but the energy barrier between A and I_1 would be smaller, so the reaction A $\Rightarrow$ B would proceed much faster.

(3.1.2)

a) The reaction is spontaneous to the left. Therefore, the hydrolysis of glucose 6-phosphate is the spontaneous reaction (glucose 6-phosphate $\Rightarrow$ glucose + P_i).

b) Since $\Delta G'_0$ is < 0, the reaction is spontaneous to the right. Therefore, the hydrolysis of ATP is the spontaneous reaction (ATP $\Rightarrow$ ADP + P_i).

c)

	glucose + P_i $\Rightarrow$ glucose 6-phosphate + H_2O	$\Delta G'_0$ = +3.3 kcal/mol
+	ATP + H_2O $\Rightarrow$ ADP + P_i + H^+	$\Delta G'_0$ = –7.3 kcal/mol
	glucose + ATP $\Rightarrow$ glucose 6-phosphate + ADP + H^+	$\Delta G'_0$ = –4.0 kcal/mol

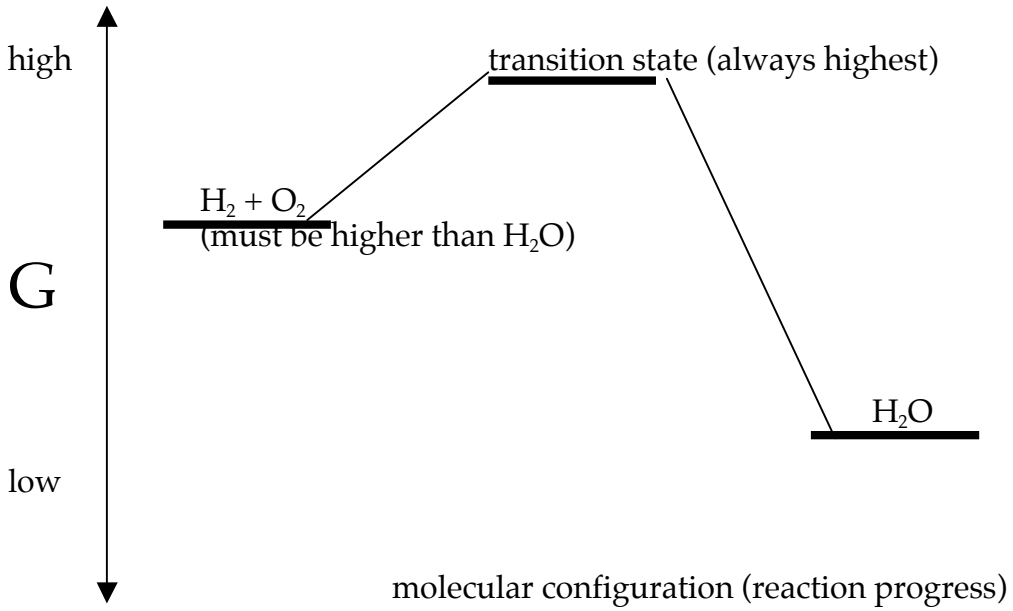
Since $\Delta G'_0$ is < 0, the overall reaction is spontaneous to the right. The nonspontaneous formation of glucose 6-phosphate is coupled to the very spontaneous hydrolysis of ATP, resulting in a net spontaneous formation of glucose 6-phosphate from glucose and ATP.

(3.1.3)

a) The highest energy species are the nonbonded atoms; these are state 2. The two bonded states will have lower energies. Since the reaction from H_2 + O_2 to H_2O is spontaneous, we know that the products (H_2O) will have lower energy than the reactants (H_2 + O_2). Therefore, from highest energy to lowest energy, it will be 2, 3, then 1.

b) Covalent bonds are broken when you go from state 3 to state 2. You are breaking the molecules into atoms (covalent bonds hold atoms together to form molecules).

c) The diagram would look like this (only the relative levels are important here, not the actual values):



(3.1.4)
a) The water slows the rate at which atmospheric O_2 gets to the enzyme, thus slowing the reaction. When the potatoes are again exposed to air, the enzyme and all substrates are present at high levels, so the reaction proceeds rapidly.

b) The cooking destroys the enzyme. The heating causes the enzyme to denature (the three-dimensional structure is lost) and thus to become inactive.

c) The enzyme is not active at the lower pH.

(3.1.5)
a) Penicillin inactivates transpeptidase, making it impossible for the bacteria to synthesize their cell wall. Without a cell wall, the bacterial membrane is not supported, and it can burst, killing the cell.

b) Human cells do not have a cell wall; therefore, they do not need transpeptidase, so penicillin has no effect on them.

c) The β-lactamase must destroy the penicillin before it can inactivate the transpeptidase. β-Lactamase is therefore likely to be located in the cell wall, outside of the cell membrane (in real life, it is).

(3.1.6)
The targets for parathion and paraoxon are virtually identical in insects and humans. Parathion itself is relatively inert, but paraoxon is very reactive toward both the human and insect target molecules. Paraoxon acts by covalently binding to and irreversibly inhibiting a critical enzyme in insects and humans. In the situation we described, humans are exposed only to parathion, which is not very toxic. Insects convert the parathion to paraoxon, which is very toxic. The insect's ability to convert the parathion to paraoxon makes parathion more toxic to insects than humans.

(3.1.7)
a)

i) By inhibiting COX-2, aspirin prevents production of prostaglandin E_2. The resulting lower levels of prostaglandin E_2 cause the pain-sensitive cells to send fewer and weaker pain messages to the brain. This reduces the sensation of pain.

ii) By inhibiting COX-1, aspirin prevents production of prostaglandin E_1. The resulting lower levels of prostaglandin E_1 reduce the production of new stomach lining cells. When these cells are not replaced, the stomach lining can break down, leading to an ulcer.

b) COX-2 inhibitors have all the analgesic properties of aspirin, but since they don't inhibit COX-1, the stomach lining is unaffected and the risk of ulcers is reduced.

(3.2) Biochemical pathways, general

(3.2.1)

$$\text{compound X} \xrightarrow{\text{enzyme 1}} \text{compound Y} \xrightarrow{\text{enzyme 2}} \text{compound Z} \xrightarrow{\text{enzyme 3}} \text{compound A}$$

Bacterial cells with defective enzymes 1, 2, or 3 will grow only if compound A is available (added to the growth media).

a) Each enzyme acts on a substrate to create a new compound. If an enzyme is nonfunctional, then the substrate cannot be used to create a new compound, and the substrate may accumulate. Compound X will accumulate in the cells defective in enzyme 1, compound Y will accumulate in the cells defective in enzyme 2, and compound Z will accumulate in the cells defective in enzyme 3.

b) Assume that for the cell to live it must have a supply of compound A. A cell lacking enzyme 1, 2, or 3 will not survive because compound A is not made. If a cell lacking only enzyme 1 is provided with a usable source of compound Y, then compound Y can be converted to compound Z and compound Z can be converted to compound A and the cell will survive. Likewise, a cell lacking only enzyme 1 can be provided with a usable source of compound Z or compound A and survive. However, if a cell lacking only enzyme 2 is provided with a usable source of compound Y, it will not survive because compound Y cannot be converted to compound Z. If a cell lacking only enzyme 2 is provided with a usable source of compound Z or compound A, it can survive. A cell lacking enzyme 3 must be provided with a usable source of compound A.

c) The compound that is acted on by the first enzyme in the pathway will be the compound that builds up. So in a cell lacking enzyme 1 and enzyme 2, compound X will accumulate. A cell lacking enzyme 2 and enzyme 3 will accumulate compound Y, and a cell lacking enzyme 1 and enzyme 3 will accumulate compound X.

d) Only compound(s) that can be acted on by functional enzymes to make compound A will allow the cells to survive. A cell lacking enzyme 1 and enzyme 2 can survive only if supplied with compound Z or compound A. A cell lacking enzyme 2 and enzyme 3 can survive only if supplied with compound A. Likewise, a cell lacking enzyme 1 and enzyme 3 can survive only if supplied with compound A.

(3.2.2)
Remember that arginine is the ending point and must be produced if the cells are to live. The pathway has at least four enzymes. Therefore, even before examining the data in the table, you can sketch out the pathway:

compound __ $\overset{\text{enzyme}}{\Rightarrow}$ compound __ $\overset{\text{enzyme}}{\Rightarrow}$ compound __ $\overset{\text{enzyme}}{\Rightarrow}$ compound __ $\overset{\text{enzyme}}{\Rightarrow}$ arginine

From the table we know that cells missing enzyme 2 survive only if arginine is added. So we can draw the following:

compounds A, B, C $\overset{\text{enzyme 2}}{\Rightarrow}$ arginine Thus, enzyme 2 is the last enzyme in the pathway.

From the table we know that cells missing enzyme 1 survive (make arginine) if compound B is added. Therefore, compound B must be later in the pathway than compounds A or C. Building on the previous step, we can draw:

compounds A, C $\overset{\text{enzyme 1}}{\Rightarrow}$ compound B $\overset{\text{enzyme 2}}{\Rightarrow}$ arginine

From the table we know that cells missing enzyme 3 survive (make arginine) if compound A or B is added. Therefore, compound A must be later in the pathway than compound C. Building on the previous step, we can draw:

compound C $\overset{\text{enzyme 3}}{\Rightarrow}$ compound A $\overset{\text{enzyme 1}}{\Rightarrow}$ compound B $\overset{\text{enzyme 2}}{\Rightarrow}$ arginine

Thus, the overall pathway is:

$\overset{\text{enzyme 4}}{\Rightarrow}$ compound C $\overset{\text{enzyme 3}}{\Rightarrow}$ compound A $\overset{\text{enzyme 1}}{\Rightarrow}$ compound B $\overset{\text{enzyme 2}}{\Rightarrow}$ arginine

(3.2.3)

Enzyme	**Inhibitors**	**Explanation**
a	chorismate	If this were inhibited by Phe, Trp, or Tyr, high levels of any one could cut off the synthesis of all three, which would be bad for the cell. The only appropriate condition when this enzyme should be inhibited is where ALL three amino acids are in excess. Chorismate will build up only when all three amino acids are in excess [because later steps (**b** and **c**) will be inhibited], so chorismate is the appropriate regulator.
b	prephenate	Similar reasoning as for **a** (this is also product inhibition).
c	tryptophan	It is the final product of the dedicated pathway where Trp is the final product.
d	phenylalanine	Same reasoning as for **c**.
e	tyrosine	Same reasoning as for **c**.

Note: All enzymes are also inhibited by accumulation of their immediate products by mass action (e.g., **d** by phenylpyruvate). This is not allosteric feedback regulation, but product inhibition.

(3.3) Glycolysis, respiration, and photosynthesis

(3.3.1)

a) This is an example of a phosphoryl transfer (phosphorylation) reaction. It is similar to the reaction catalyzed by **hexokinase,** which converts glucose into glucose 6-phosphate. It is also similar to the **phosphofructokinase** reaction, which converts fructose 6-phosphate into fructose 1,6-bisphosphate. Also, it is similar to the **phosphoglycerate kinase** reaction, which converts 1,3-BPG into 3PG, making ATP in the process. (Note that an enzyme can catalyze both the forward and reverse reactions at each step.) It is not similar to the pyruvate kinase reaction, which uses a reactive dehydrated C=C bond to drive the synthesis of ATP in converting PEP to pyruvate.

b) This is an isomerization reaction. It is similar to the one catalyzed by **phosphoglucoisomerase**, which converts glucose 6-phosphate into fructose 6-phosphate. It is also similar to the **triose phosphate isomerase** reaction, which converts dihydroxyacetone phosphate into glyceraldehyde 3-phosphate.

c) Each enzyme specifically recognizes and binds to the correct substrate. Not all of the above substrates would fit correctly in the right orientation in the enzyme's substrate-binding region.

d) Malonate is very similar in structure to the normal substrate, succinate. The most likely possibility is that it is competing with succinate for the enzyme's substrate-binding region.

(3.3.2)

a) In the presence of O_2, pyruvate is no longer converted to lactic acid, but instead is converted to acetyl-CoA, which can enter the citric acid cycle. Thus, glucose is converted to CO_2 and H_2O.

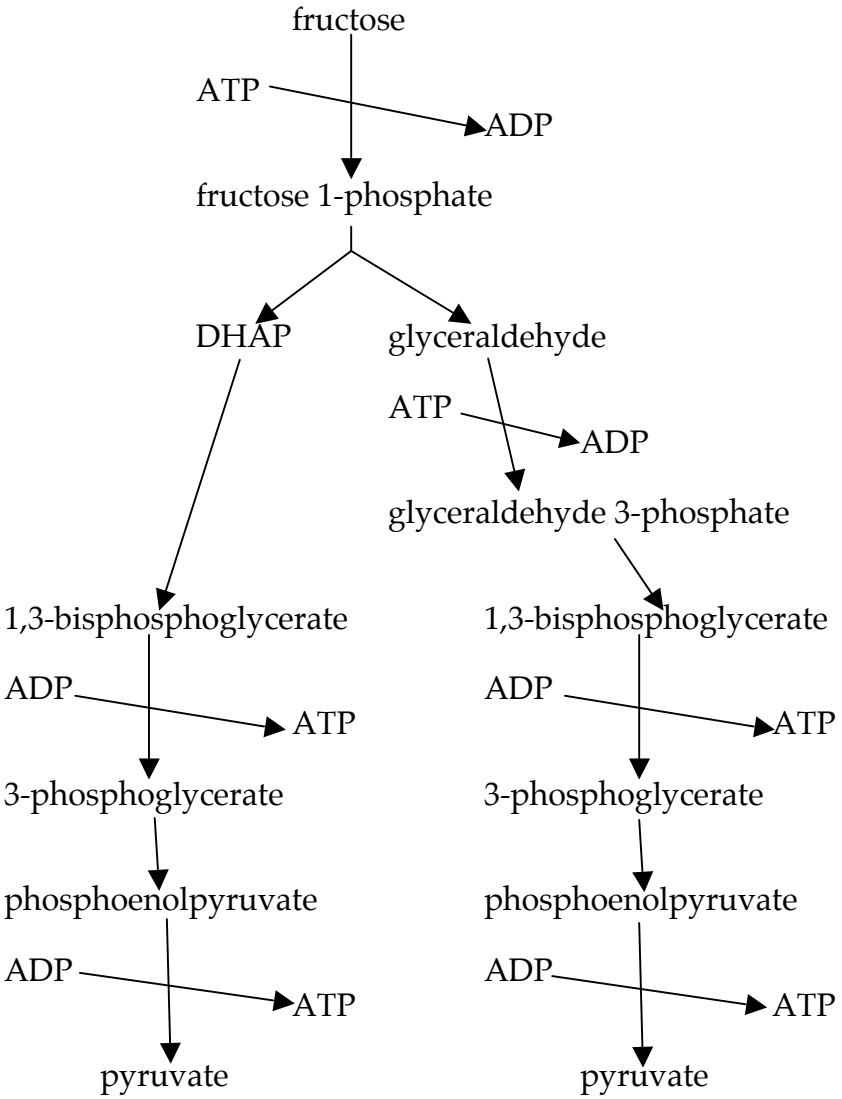
b) Growth requires energy and a cell gets roughly 36 ATP/glucose with O_2 around, as opposed to 2 ATP/glucose without O_2. Therefore, cells require less glucose for the same growth rate (ATP consumption rate) in the presence of O_2.

c) NAD^+ is a cofactor, while glucose is a substrate. Thus, glucose is consumed in large amounts and not recycled. NAD^+ is used to carry the electrons removed from glucose and its derivatives (becoming NADH), but it is recycled to NAD^+ by transferring electrons to other substrate molecules: under anaerobic conditions, to pyruvate; under aerobic conditions, to O_2 via the electron transport chain. Since NAD^+ is recycled, only a small amount is necessary to process a large amount of glucose.

(3.3.3)
a) ATP is the most likely source of phosphate and energy for this reaction. By analogy to reaction 1 it would most likely be:

glyceraldehyde + ATP ⇒ glyceraldehyde 3-phosphate + ADP

b)



c)

	Reaction	ATP used or produced per fructose
Fructose utilization	fructose to fructose-1-P	1 used
	glyceraldehyde to glyceraldehyde-3-P	1 used
	two 1,3-bisphosphoglycerates to two 3-phosphoglycerates	2 produced
	two phosphoenolpyruvates to two pyruvates	2 produced

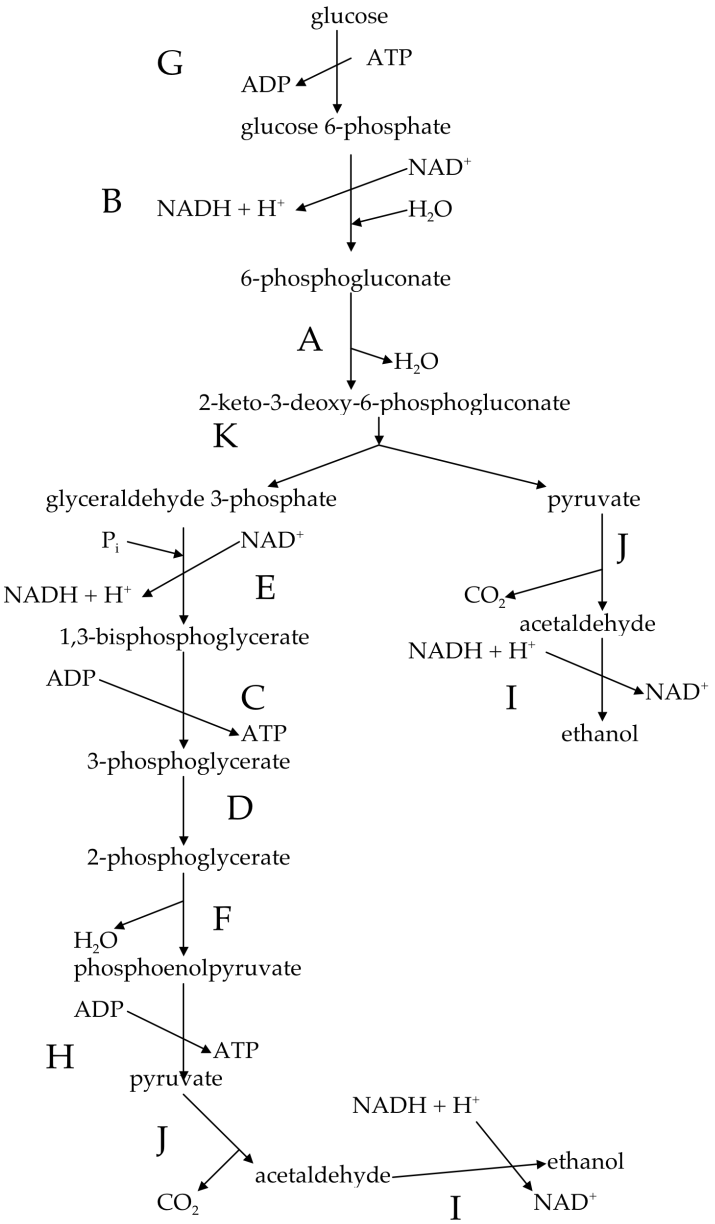
Net: +2

	Reaction	ATP used or produced per fructose
Glucose utilization	glucose to glucose-6-P	1 used
	fructose-6-P to fructose 1,6-bisphosphate	1 used
	two 1,3 -bisphosphoglycerates to two 3-phosphoglycerates	2 produced
	two phosphoenolpyruvates to two pyruvates	2 produced

Net: +2

(3.3.4)
a)

(3.3.4)
a)



40 of 43

b) The overall reaction is:

$$\text{glucose} + \text{ADP} + P_i \longrightarrow 2CO_2 + 2\ \text{ethanol} + \text{ATP} + H_2O$$

Therefore, the bacterium gets only 1 ATP per glucose as compared with 2 ATP/glucose from the Embden-Meyerhof-Parnas pathway.

c) Consider the fate of the two glucose 6-phosphate and glyceraldehyde 3-phosphate produced from three xyloses:

Each glucose 6-phosphate will enter at the second step of the Entner-Doudoroff pathway, giving the following reaction:

$$\text{glucose 6-phosphate} + 2\ \text{ADP} + P_i \longrightarrow 2\ CO_2 + 2\ \text{ethanol} + H_2O + 2\ \text{ATP}$$

The glyceraldehyde 3-phosphate will enter at the fifth step of the Entner-Doudoroff pathway, giving the following reaction:

$$\text{glyceraldehyde 3-phosphate} + 2\ \text{ADP} + P_i \longrightarrow CO_2 + \text{ethanol} + H_2O + 2\ \text{ATP}$$

d) Adding the three reactions:

	3 xylose + 3ATP	⇒	2 glucose 6-phosphate + glyceraldehyde 3-phosphate (G3P) + 3 ADP
+	2 glucose 6-phosphate + 4 ADP + 2 P_i	⇒	4 CO_2 + 4 ethanol + 2 H_2O + 4 ATP
+	glyceraldehyde 3-phosphate + 2 ADP + P_i	⇒	CO_2 + ethanol + H_2O + 2 ATP
	3 xylose + 3 ADP + 3 P_i	⇒	5 CO_2 + 5 ethanol + 3 H_2O + 3 ATP

e) From the above, 3 ATP for 3 xylose = 1 ATP per xylose.

(3.3.5)

a) The biochemical process responsible for the plant's absorption of O_2 and production of CO_2 in the dark is respiration. The plant cells are using O_2 to oxidize stored carbohydrates to CO_2, to produce energy.

b) Yes, this process continues when the plant is exposed to light. The plant always needs energy from respiration to perform cellular reactions. In the light, the rate of photosynthesis is greater than the rate of respiration, so the result is net O_2 production and net CO_2 absorption. In the dark, there is no photosynthesis, so the basal respiration level predominates (O_2 absorption and CO_2 production).

(3.3.6)

a and b) Since NADPH will no longer be produced, the dark reactions (Calvin-Benson cycle) will stop. As a result, these reactions will no longer consume ATP, causing ATP to build up and ADP to disappear. Without ADP, the H^+ATPase will not be able to let

H^+ flow down the concentration gradient and the H^+ gradient will build up. When the H^+ gradient gets large enough, e^- transport will stop because there won't be enough energy to pump H^+ against the gradient. If this happens, O_2 will no longer be produced. Since it will take time for all this to back up, O_2 will be produced for a while, then stop.

c)

i) No. If e^- can no longer be transferred to pQ, e^- will back up in pheophytin I (sometimes called simply "I" in various textbooks) and e^- transport will come to a halt. Without a place for the e^- from H_2O to go, H_2O will no longer be oxidized to O_2.

ii) No. With no e^- flowing down the transport chain, no H^+ will be translocated across the thylakoid membrane, no H^+ gradient will form, and therefore no ATP can be made.

iii) No. This process requires reducing power in the form of NADPH. Since there are no e^- flowing through the transport chain, there won't be any to reduce $NADP^+$, so there will be no NADPH around to reduce the CO_2.

(3.3.7)

a)

Enzyme	Pathway
lactate dehydrogenase	This is the last step in lactic acid fermentation.
alcohol dehydrogenase	This is the last step in alcohol fermentation.
aldolase	This is a key step in glycolysis.
proton ATPase	This is a key enzyme in oxidative phosphorylation.
cytochrome *c*	This is part of electron transport.
citrate synthase	This is the "first" step in the citric acid cycle.

b) Based on the above, you can rewrite the table as follows:

Organism	Lactic acid fermention	Alcohol fermentation	Glycolysis	Oxidative phosph.	Electron transport	Citric acid cycle
Haemophilus influenzae	No	Yes	Yes	Yes	No	No
Mycoplasma genitalium	Yes	No	Yes	Yes	No	No
Mycoplasma pneumoniae	Yes	No	Yes	Yes	Yes	No
Escherichia coli	No	Yes	Yes	Yes	Yes	Yes

Thus,

– in the presence of oxygen:

Haemophilus influenzae: Although this organism has the proton ATPase, it lacks the other components of cellular respiration. It must therefore carry out fermentation. In this case, it would produce alcohol + CO_2 and get about 2 ATP per glucose.

Mycoplasma genitalium: This lacks the capability for electron transport or the citric acid cycle; so, even though it has the proton ATPase, it cannot carry out cellular respiration. It can still ferment glucose to lactic acid, so it will produce lactic acid and roughly 2 ATP per glucose.

Mycoplasma pneumoniae: This also cannot perform cellular respiration, so it must ferment glucose. It has the enzymes to ferment glucose to lactic acid and will get roughly 2 ATP per glucose.

Escherichia coli: This has all the parts needed for cellular respiration so it will make CO_2 + H_2O and get roughly 36 ATP per glucose.

– in the absence of oxygen, the first three bacteria will perform the same fermentations. *Escherichia coli* will ferment glucose to ethanol and CO_2 and get roughly 2 ATP per glucose.

Molecular Biology Solutions

(1) PROBLEMS EXPLORING CLASSIC EXPERIMENTS

(1.1)
a) The negative control shows that either $R_{(II)}$ or heat-killed S_{III} alone is nonvirulent. Without this result, you could argue that one or the other killed the mice, and it would not be necessary to invoke transformation to explain the results of the central experiment.

If the negative control experiment resulted in dead mice, the result of the central experiment would have been meaningless.

The positive control shows that the bacteria do not change type once inside the mouse. Without this result, you could argue that the bacteria randomly change type. It also shows that the bacteria are still lethal to the mouse. This control shows more directly that S_{III} is the cause of death. Without this result, you could argue that the mouse died of some other cause and just happened to harbor S_{III} bacteria.

If the positive control experiment resulted in mice that lived, you could not have done the central experiment because there would be no difference between injecting the mixture or injecting either $R_{(II)}$ or heat-killed S_{III} alone.

b) It was essential to use a mixture of bacteria that are derived from different strains, i.e., $R_{(II)}$ and heat-killed S_{III} instead of a mixture of $R_{(II)}$ and heat-killed S_{II}. This excludes the possibility that a few $R_{(II)}$ bacteria had reverted back to S_{II} and these S_{II} were responsible for the death of the mice. In this case, it would not be necessary to invoke transformation to explain the results of the central experiment. However, because $R_{(II)}$ and S_{III} are derived from different strains, there can be no reversion. So when $R_{(II)}$ and heat-killed S_{III} were injected, and virulent S_{III} are found, we are forced to consider transformation to explain the results of the central experiment.

(1.2)
a) They presumed that the transforming substance was genetic material, because the transformation from R to S was heritable. Once an R cell was transformed into an S cell, all the progeny from that cell and successive generations were S cells.

b) Even the smallest trace of protein in their preparations of the transforming substance could support model (1). It could be that the DNA is simply a carrier for the proteins that are the true genetic material. In that case, you might need only a very little protein to make a gene.

c) i) In the four base pairs shown, there are 30 N's and 8 P's. To get the mass ratio, you must multiply by the atomic weights of each atom (N = 14, P = 31):

$$\frac{30 \times 14}{8 \times 31} = 1.69$$

ii) Protein contamination should increase the N/P ratio. Protein contains N, but no P.

iii) These data (with the exception of #38B) show a ratio of 1.69 or below. So it appears that the preparations were not contaminated with protein. However, it is not clear why two of the N/P ratios are ≤ 1.69.

d) i) These data support the model that genes are made of DNA and not protein. If the genetic material were protein, then treatment with these enzymes should destroy its activity. If it were DNA, they should have no effect.

ii) This is a negative result; there was no effect. Because of this, you can never be sure that the enzymes were working properly or if somehow the proteins that make up genes are resistant to these enzymes.

While this work was compelling, it was not completely conclusive. Avery, McCarthy, and MacLeod did several additional experiments. The accumulation of data, all consistent with model (2), gradually convinced the scientific community. This is typical of the process of science; there is almost never one crucial experiment that answers a question once and for all.

(1.3)

a) i) If the phage injected protein, not DNA, then in experiment 1 you would expect to find most of the ^{32}P in the supernatant because the ^{32}P-labeled DNA was never injected into the bacterium.

If the phage injected protein, not DNA, then in experiment 2 you would expect some ^{35}S in both supernatant (sheared-off phage heads) and the pellet (transferred genetic material). The progeny phage would have some ^{35}S from the reused genetic material.

ii) If the phage injected genetic material that was mostly DNA and a little protein, then in experiment 1 you would find little ^{32}P in the supernatant and a lot in the pellet since it has been injected into the bacteria.

If the phage injected genetic material that was mostly DNA and a little protein, then in experiment 2 you would expect ^{35}S in both supernatant (sheared-off phage heads) and a little in the pellet (transferred genetic material). The progeny phage would have some ^{35}S from the reused genetic material.

iii) If the phage injected DNA and protein, but protein is the genetic material, and DNA is only a scaffold, then in experiment 1 you would find little ^{32}P in the supernatant and a lot in the pellet since it has been injected into the bacteria.
If protein is the genetic material, and DNA is only a scaffold, then in experiment 2 you would expect some ^{35}S in both supernatant (sheared-off phage heads) and in the pellet (transferred genetic material). The progeny phage would have some ^{35}S from the re-used genetic material.

b) i) ^{32}P in the supernatant represents phage DNA that was not injected into the bacteria (perhaps the phages were sheared off before they injected their DNA) or DNA present in phage that did not attach to the cells in the first place. It is not crucial that

this number be zero because their model did not require that all the DNA of all the phages had to be injected.

ii) ^{35}S in the pellet represents protein that remained associated with the bacteria. Perhaps not all phages are sheared off the blender, or perhaps a small portion of the phage is left attached to the bacteria. ^{35}S in the pellet makes the experiments less conclusive, since it could also represent protein that was injected along with the DNA, which could be genetic material or other proteins. This undermines the model that only DNA injection is required for infection and that therefore DNA is the genetic material.

c) You can rule out only the possibility that the genetic material injected was protein and the DNA remained in the phage head (model 1). This model predicts that very little ^{32}P would go into the bacteria, but that is not what the data show. Both other models are consistent with the data.

(1.4)
a)

- Nucleic acid will be labeled with ^{32}P?
- All macromolecules will be labeled with ^{3}H?
- Proteins will be labeled with ^{35}S?

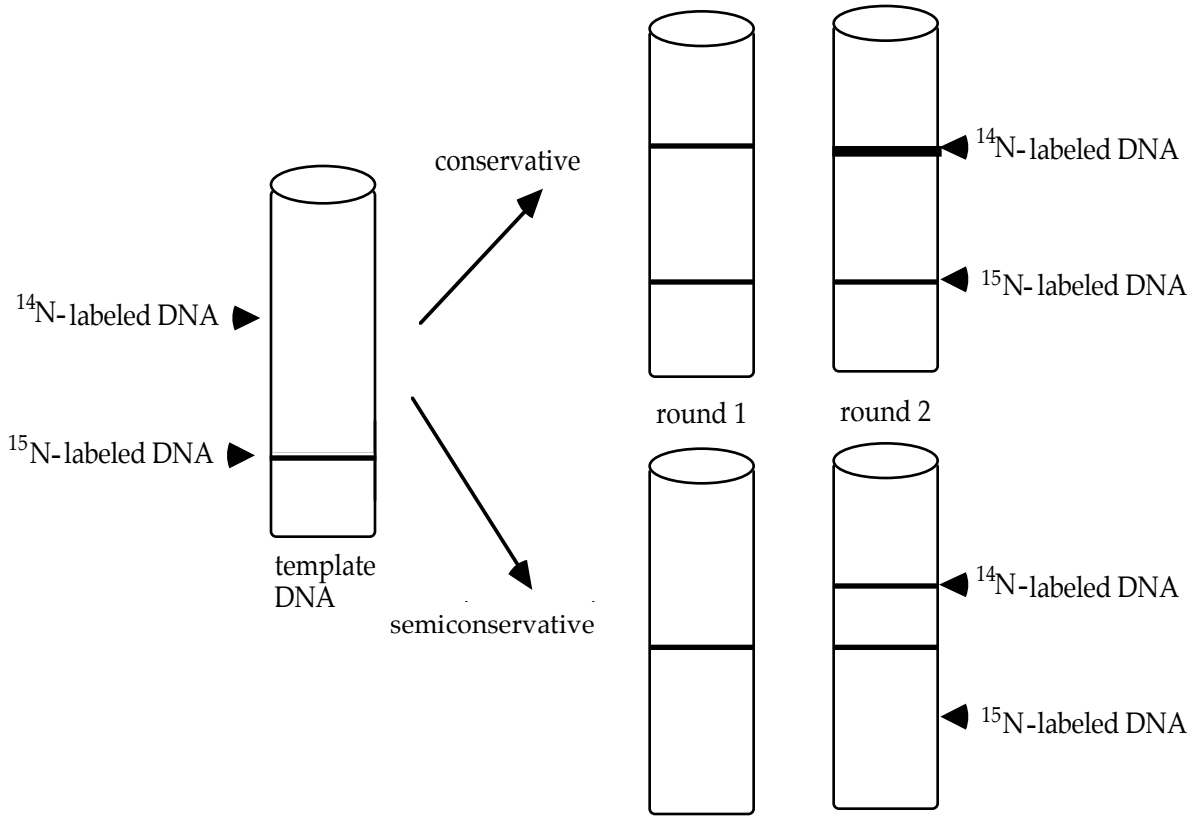
b) This virus is carrying double-stranded DNA as shown by the %A = %T, %C = %G. The trace of uracil is contaminating RNA.

c) In a manner analogous to the Hershey-Chase experiment, you would infect cells with ^{32}P labeled virus. After a short time, you would separate the virus particles from the cells and examine whether the ^{32}P-labeled DNA is found in the virus particles (in the supernatant) or in the bacteria (in the pellet).

d) i) An in vitro system for DNA replication would include template DNA, primer, dNTPs, DNA polymerase, ligase, helicase, topoisomerase, single-stranded binding protein.

ii) Nitrogen is found in the bases of DNA.

iii) You repeat the Meselson-Stahl experiments. On the diagram below, draw the results expected at each round for both conservative and semiconservative replication.



(2) PROBLEMS EXPLORING THE STRUCTURE OF DNA AND RNA

(Computer Activity 1)

a) i) The strands are antiparallel. They run 5′ to 3′ in opposite directions. If you reset the view and click the (i) button:

- The strand that starts at the top left (light yellow) runs 5′ to 3′ top to bottom.
- The strand that starts at the top right (light purple) runs 3′ to 5′ top to bottom.

ii) There are no covalent bonds between the bases, only hydrogen bonds. Therefore, this view shows two molecules of DNA and the entire object should be called "a double-stranded DNA complex" ("double-stranded DNA molecule" is, strictly speaking, incorrect). These are two DNA strands.

In the DNA double helix, the bases are on the inside of the helix; in a protein alpha-helix, the side chains are on the outside.

iii) A always pairs with T and G always pairs with C. The complete DNA sequence is as follows: the top strand corresponds to the strand that starts with 5′ at the upper left of the starting display. The | correspond to the hydrogen bonds in the base pairs.

```
5′-CTTCCTCATGTATATACATGAGGA-3′
   |||||||||||||||||||||||||
3′-GAAGGAGTACATATATGTACTCCT-5′
```

b) i) In the reset view, the left strand runs 5′ to 3′ top to bottom and the right strand runs 3′ to 5′ top to bottom.

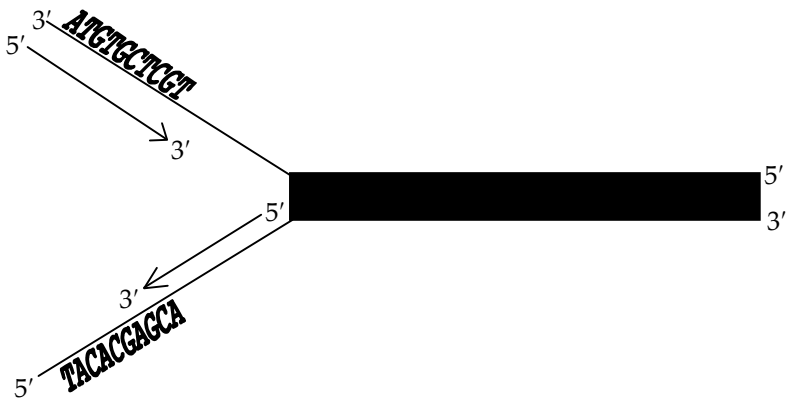
ii) The top strand corresponds to the left strand in the reset display:

```
5′-TCC-3′
   |||
3′-AGG-5′
```

(2.2)

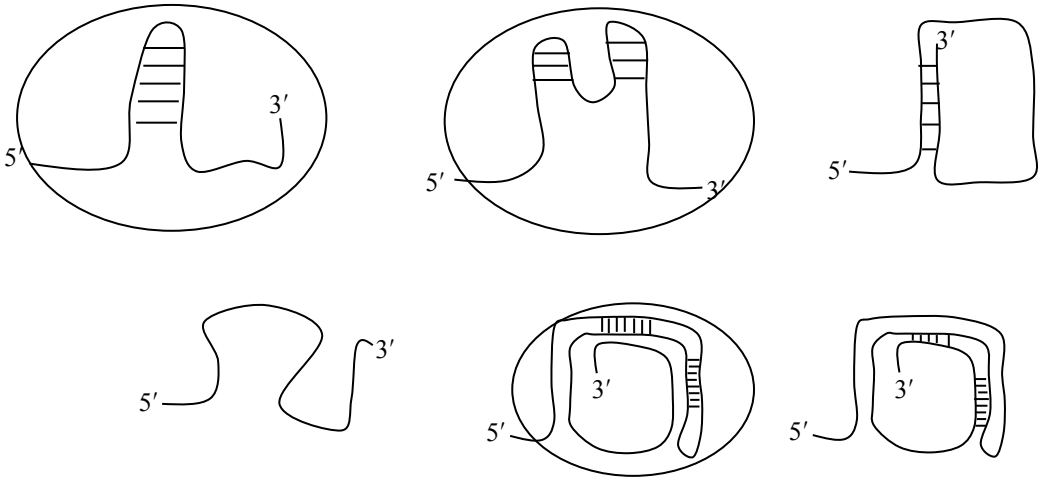
a) The sequence of the newly formed DNA is 5′-TACACGAGCA-3′.

b and c)



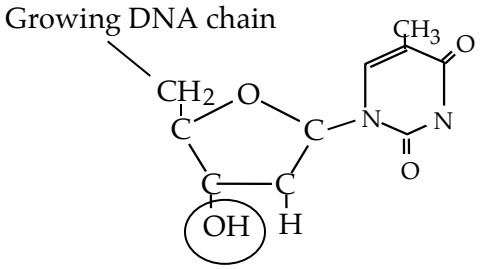
(2.3)

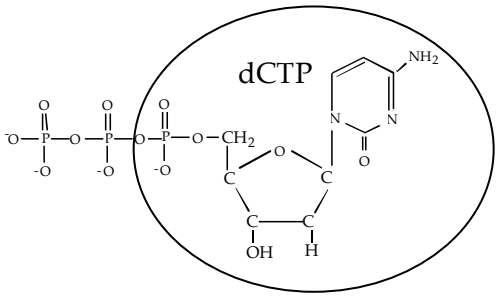
a)



b) 3′ …TGCCTGCG…5′

(2.4)





(3) DNA REPLICATION

(3.1)

Enzyme activity	Function(s)
Topoisomerase	k
Primase (synthesizes primer)	e, f, j
DNA polymerase to elongate new DNA strand	b, c
Helicase to unwind DNA	h
DNA polymerase to replace RNA with DNA	d
Processivity factor	i

Choose from:

a) 3′ ⇒ 5′ growth of new DNA strand

b) 5′ ⇒ 3′ growth of new DNA strand

c) 3′ ⇒ 5′ exonuclease

d) 5′ ⇒ 3′ exonuclease

e) Makes RNA primer complementary to the lagging strand

f) Makes RNA primer complementary to the leading strand

g) Makes peptide bonds

h) Separates the two DNA strands

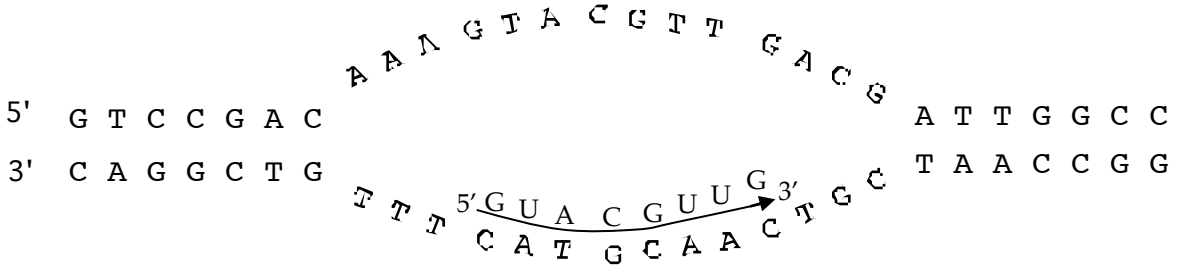
i) Maintains DNA polymerase on template

j) Provides 3′-hydroxyl for initiation of DNA polymerization

k) Untangles super-coiled DNA

(3.2)

a)



b) If the replication fork moves to the left, this primer will be used to create the lagging strand because the direction of replication will be moving in the opposite direction of the movement of the origin of replication.

c) Replication is discontinuous because the origin is moving 3′ to 5′ relative to the lagging strand, but nucleotides can be added only in the 5′ to 3′ direction. On the lagging strand, each time a primer is created the replication moves in the opposite direction of the origin. Thus, as the replication fork moves and the DNA is unwound, more unreplicated DNA at the 5′ end of the primer is revealed. Since DNA can only be replicated 5′ to 3′, new primers must constantly be added, making a discontinuous strand of DNA.

(3.3)

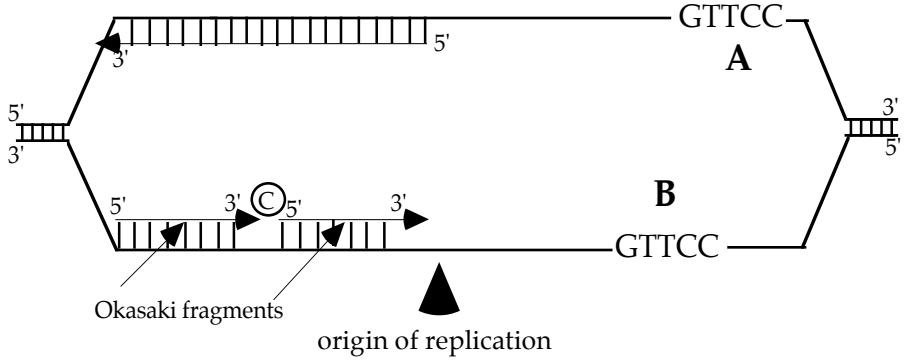
a) i) Replication will be continuous on templates 1 and 4

ii) The primer 5′-GUUCC-3′ binds to and initiates replication at sites B and C.

iii) The lagging strand is more affected by the lack of DNA ligase. DNA replication on the lagging strand occurs in small stretches called Okasaki fragments. For replication of the lagging strand to be complete, a phosphodiester bond must be formed between the 3′-OH on one Okasaki fragment and the 5′-phosphate on the other. DNA ligase makes this bond.

(3.4)

a)



b) DNA ligase is required at C in the above diagram.

c) The primer 5′-CAAGG-3′ binds at site B to initiate replication.

d) From site B, the direction of elongation of the daughter DNA strand is to the right.

e) From site B, DNA synthesis is performed in a continuous fashion relative to the nearest replication fork.

(3.5)

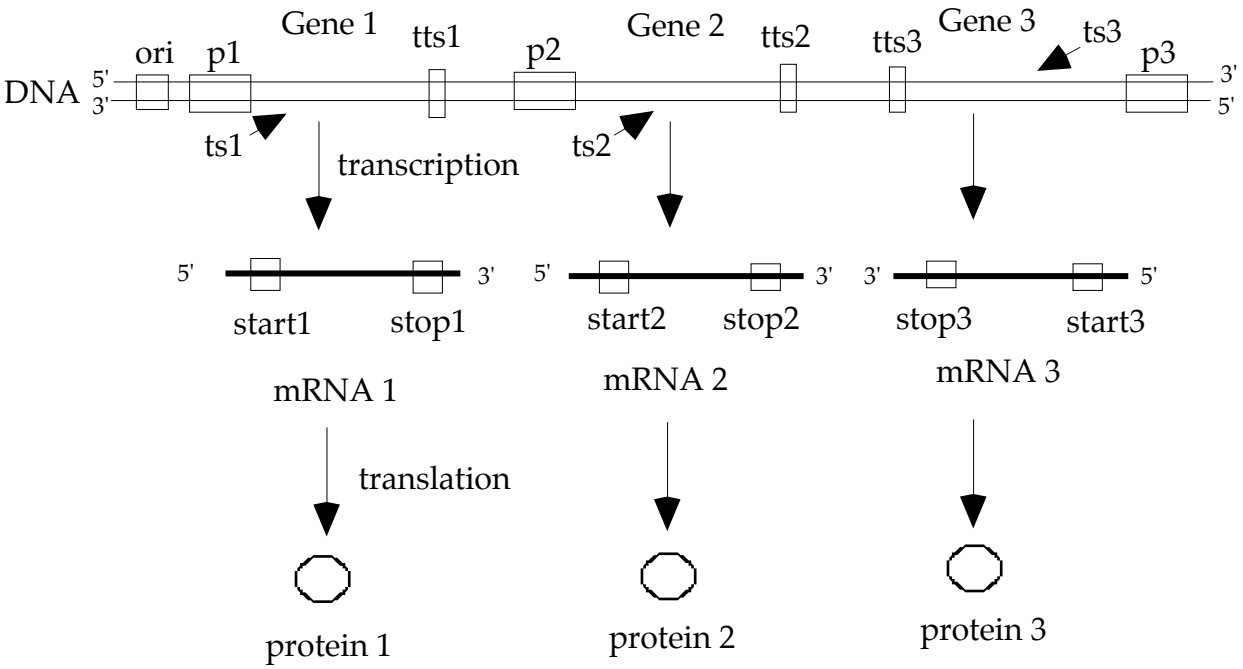
a) The sequence of the RNA primer that binds to the top strand at base-pair positions 80–90 is 5′-UGUACGCAUGC-3′.

b) DNA synthesis from the primer in (a) would be continuous to the left as the diagram is displayed.

c) The sequence of the RNA primer that binds to the bottom strand at base-pair positions 90–100 is 5′-AUAGUUCGACG-3′.

d) DNA synthesis from the primer in (c) above would be continuous to the right as the diagram is displayed.

(4) TRANSCRIPTION AND TRANSLATION
(4.1) Transcription and translation in prokaryotes
(4.1.1)
a)



promoter - DNA sequence recognized by RNA polymerase. Signals RNA polymerase to start transcription; always found at the 5′ end of the transcribed region.
transcription termination site - DNA sequence recognized by RNA polymerase. Signals RNA polymerase to stop transcribing. This determines the 3′ end of the message.
start codon - RNA sequence recognized by ribosome. Signals ribosome to begin translation; often found near the 5′ end of mRNA.
stop codon - RNA sequence recognized by ribosome. Signals ribosome to end translation; often found near the 3′ end of mRNA.
transcribed strand - DNA strand that RNA pairs with during transcription; has orientation opposite to mRNA produced.

b) No. If RNA polymerase II can't be recruited to the promoter, the mRNA for gene 2 cannot be transcribed, so no protein will be produced.

c) Because the transcription termination site in gene 1 was mutated, RNA polymerase II can continue making an mRNA molecule until it encounters the next available transcription termination site. Apparently, this was the one located in gene 2. Thus, one long mRNA contains the protein-coding regions for both gene 1 and gene 2. Each of these protein-coding regions is preceded by a start codon, so each will be translated by ribosomes as they scan the mRNA.

(4.1.2)
a) 5′ AAUUGUGAAU....3′

b) 5′ AAUUCCGAGC...3′

c) The mRNA has the same sequence as the DNA strand that oriented 5′ to 3′, because all nucleic acids are made in the 5′ to 3′ direction. Synthesis is directed by a template that runs antiparallel to the newly synthesized molecule. The mRNA is the same except that wherever there is a T in the DNA, there is a U in the mRNA.

d) H_3N^+-Met-Asp-Asn-Asn-Val-Thr-Gln-Glu-Thr-Ala-Lys-Thr-Met-Phe-COO^-
Protein synthesis begins at the start codon, usually the first AUG of the mRNA. Note that there can be a Met within the protein sequence; translation does not restart there. Also note that the stop codon does not specify an amino acid; the chain ends with the last amino acid before the stop codon.

e) No. Translation does not terminate there because the UAA is in a different reading frame; it is read as: GC**U** **AA**G A.

f) H_3N^+-Met-Asp-Asn-Asn-Val-Thr-Gln-Glu-Thr-Ala-Lys-Thr-Met-Phe-COO^-

For the next parts, the altered amino acids are shown in **bold and underlined** type:

g) H_3N^+-Met-Asp-Asn-Asn-**Gly-Asp-Thr-Gly-Asn-Ser**-COO^-. Note that all the amino acids after the mutation are altered.

h) H_3N^+-Met-Asp-**Lys-Met**-COO^-.

i) H_3N^+-Met-Asp-Asn-**Lys**-Val-Thr-Gln-Glu-Thr-Ala-Lys-Thr-Met-Phe-COO^-.
Note that only one amino acid is changed. This is a missense mutation.

(4.1.3)
a) 5′-AAACAGCUAUGGCCA .....-3′

b) H_3N^+-Met-Ala-Met-Ser-Thr-Pro-.....-COO^-

c) No. Translation of the mRNA does not terminate at this TAA, because nucleotides appear as AUU in the mRNA transcript. The TAA is on the noncoding strand.

d) No. Translation does not terminate there because the TAA chain-terminating codon would not be in the correct reading frame in the mRNA transcript; in the correct reading frame (set by the first AUG codon), they are read as AUU AAA.

e) H_3N^+-.....Asn-Arg-Gly-COO^-

f) A frameshift mutation as a result of the deletion would alter the sequence of the mRNA; however, the protein sequence would not be altered because the deletion is prior to the start codon.

g) There would be no change in the amino acid sequence of the protein because the codons GGC and GCA both code for the amino acid alanine.

h) The GCC in the sequence that codes for alanine would be changed to CCC, which would now code for proline.

i) The protein would be terminated prematurely because a new stop codon was created at the new nucleotide positions 55-57. As a result of this deletion, the new sequence of the protein would be: H_3N^+-Met-Gly-COO^-.

(4.1.4)

a) NH_3^+-Met-Ser-Cys-Trp-

```
     5'- AUG UCU UGU UGG ...3'
            UCA UGC
                UCG
            UCC
            AGU
            AGC
```

b) Gly is GGA, GGC, CCU, or GGG. The only way to get there via one base change is from AGU --> GGU or AGC --> GGC. So the original sequence for the Ser was AGU or AGC.

c) From (b), there are four possibilities for the normal sequence. Here they are as normal and with the first nucleotide of codon 2 deleted:

```
     (1)                  5'- AUG AGU UGU UGG...3'
with one base deleted:    5'- AUG GUU GUU GG....
                              Met Val Val Gly   NOT RIGHT

     (2)                  5'- AUG AGC UGU UGG...3'
with one base deleted:    5'- AUG GCU GUU GG....
                              Met Ala Val Gly   NOT RIGHT

     (3)                  5'- AUG AGC UGC UGG...3'
with one base deleted:    5'- AUG GCU GCU GG....
                              Met Ala Ala Gly   NOT RIGHT

     (4)                  5'- AUG AGU UGC UGG...3'
with one base deleted:    5'- AUG GUU GCU GG....
                              Met Val Ala Gly   RIGHT
```

Therefore, the original sequence was 5'-AUG AGU UGC UGG...3'.

(4.1.5)
a) No. You can model any length codon and get the same results with a homopolymer template.

b) One nucleotide per codon would only give four amino acids, so it can be ruled out. Two would work – 16 total codons could give 14 amino acids and two stops. Three is possible, but they would be degenerate or have many stop codons. To eliminate one of these two options, you can predict the results of translating $(WX)_n$ in a two or three nucleotides per codon scheme:

two nucleotides per codon could be read as WX WX WX WX or W XW XW XW X, giving a mixture of $(aa1)_m$ and $(aa2)_m$.

three nucleotides per codon could be read as WXW XWX WXW or XWX WXW XWX, giving $(aa1\text{-}aa2)_m$ only.

Since we see a mixture of two homopolymers, there must be two nucleotides per codon.

c) You can work out the code as follows:

- From the first round of experiments we know that WW = Met, XX = Val, YY = Thr, and ZZ = Leu.

- From the next round you know that either WX or XW = Ile and the other is Glu. Also, $(WY)_n$ encodes only one amino acid, so it must also encode a stop codon.

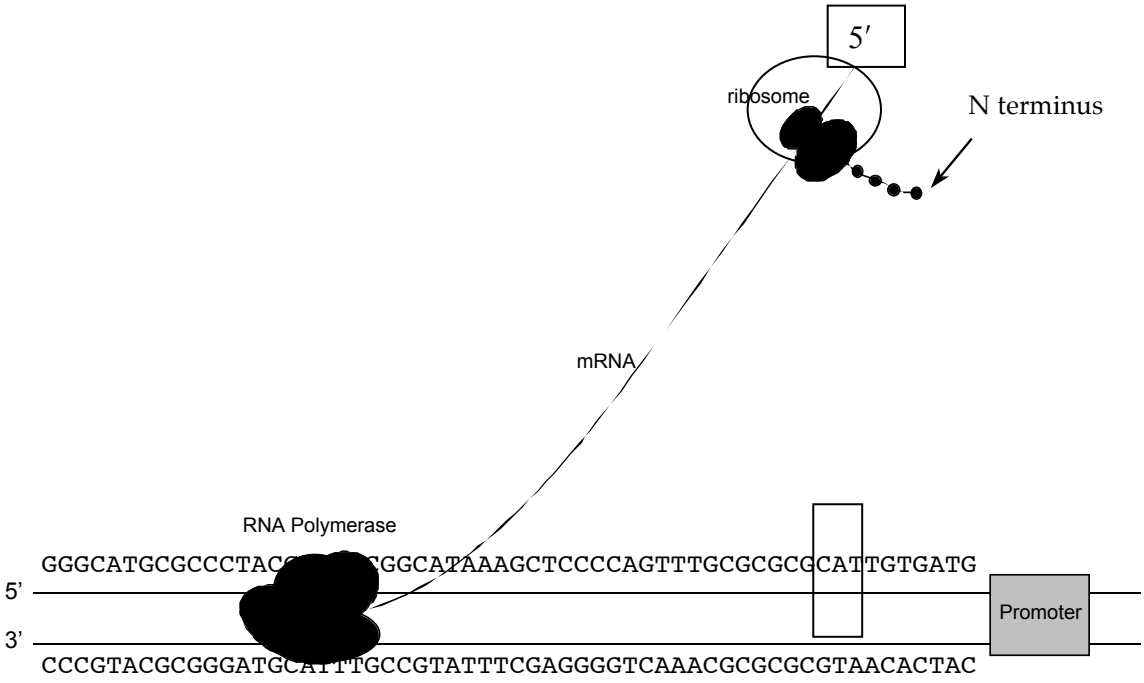
- Since from the final experiment you know that $(WXY)_n$ gives Glu but not Ile, then WX = Glu. Since $(WXY)_n$ encodes Lys, then YW = Lys, and WY = stop. By similar reasoning, the other codons can be determined:

Codon	Amino Acid
WW	Met
WX	Glu
WY	stop
WZ	Phe
XW	Ile
XX	Val
XY	Gln
XZ	Pro

Codon	Amino Acid
YW	Lys
YX	Asp
YY	Thr
YZ	Trp
ZW	Arg
ZX	Ser
ZY	stop
ZZ	Leu

(4.1.6)

a)



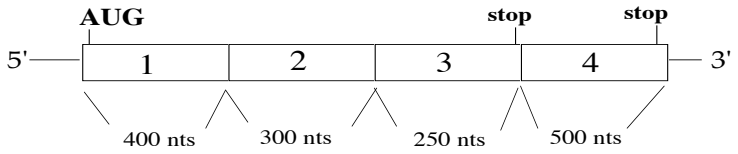
b) 3′ ACC 5′

c) Yes. Any substitution will produce a different codon, and the only codon for Trp is 5′ UGG 3′.

d) No. There are four codons for Thr. For example, 5′ ACA 3′ could be changed to 5′ ACG 3′ and still encode Thr.

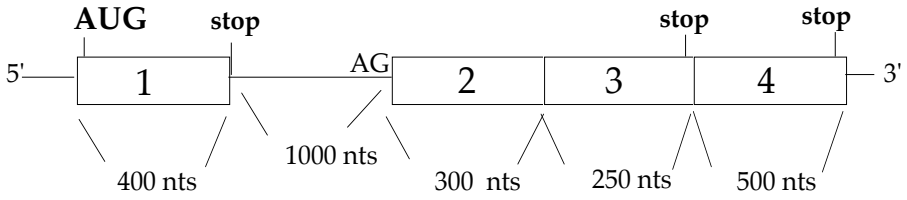
(4.2) Transcription, RNA processing, and translation in eukaryotes

(4.2.1)
a)



b) There are 400 + 300 + 250 = 950 nucleotides located between the AUG start codon and the first stop codon that is encountered. Thus, translation should result in a protein of 950/3 = 316 amino acids.

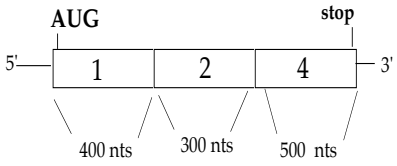
c) i)



ii) Only the 400 nucleotides of the first exon would be translated by the ribosomes before they encountered the stop codon located in intron 1, which was not removed due to the mutation. Thus, a protein of 400/3 = 133 amino acids would be expected.

iii) Since the active site, located in the C terminus of the protein, is not translated, the enzyme will not be active.

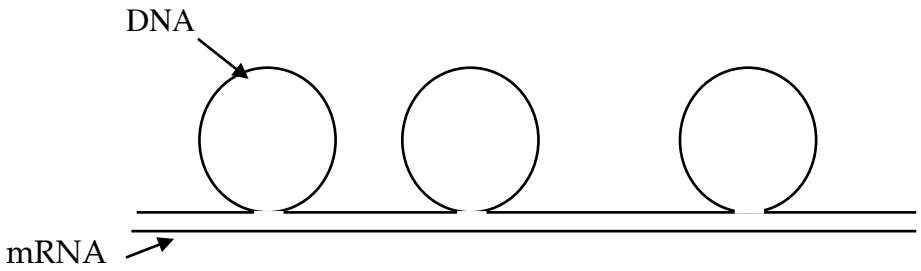
d) i)



ii) Because the 3′ splice site preceding exon 3 was altered, it could not participate in splicing. If intron 2 had simply been left in the final mRNA, then the resulting protein should be only 400 + 300/3 = 233 amino acids, owing to the stop codon present in intron 2. To get a mutant protein of 400 amino acids, 1,200 nucleotides must have been translated. This would be equivalent to the number of nucleotides present in exons 1, 2, and 4 (400 + 300 + 500 = 1,200). Thus, exon 3 is skipped entirely during the splicing process, and exon 2 is instead spliced to exon 4.

(4.2.2)

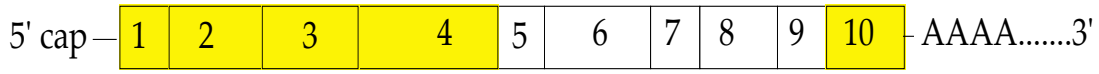
a)



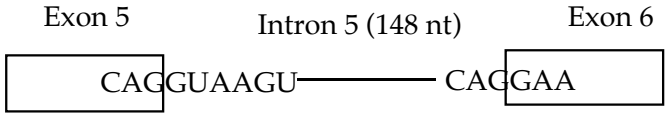
b) The genomic DNA has introns that are spliced out of the mature mRNA. Because these sequences are not represented in the mRNA, the intronic regions of the DNA can not base pair with the RNA.

(4.2.3)

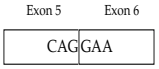
a)



b) i) before splicing:



after splicing:



ii) CAG = glutamine 206, GAA = glutamic acid 207

c) i)

ii) The mutation eliminates the 5′ splice site in intron 5. As a result, this intron will no longer be spliced out of the processed message. Three possible outcomes (you only had to give two):

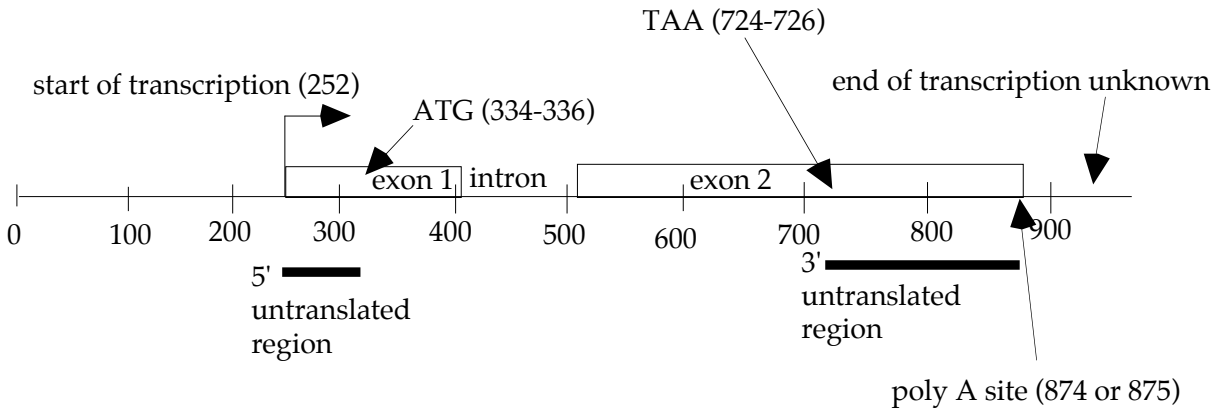
1) The unexcised intron will form a larger protein that will disrupt the protein structure.

2) The unexcised intron will introduce a stop codon that is in frame in intron 5 results in a truncated protein that is inactive since it will lack part of the NAD^+-binding domain, as well as part of the catalytic subunit.

3) The unexcised intron sequence is translated, causing a frameshift mutation (148 nucleotides = 49 codons + 1 base). The larger protein will also be inactive since it will no longer be able to properly bind NAD^+.

(4.2.4)

a) The structure of con-6 is shown below (numbers as in the genomic sequence):



b) The sequence of the con-6 protein is shown on the modified version of Fig. 2 below; spaces have been inserted to indicate the codons and the amino acid sequence is shown on an additional line:

```
GENOMIC DNA: 251 CAAACAAACACACTTCATTCCCAAGACATCACTCACAAACAACCAACCTC 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
       mRNA:   1 @AAACAAACACACUUCAUUCCCAAGACAUCACUCACAAACAACCAACCUC 49

GENOMIC DNA: 301 TTCCAATCCAACCACAAACAAAAATCAGCCAAT ATG TCC GAC TTC GAG AA 350
                 |||||||||||||||||||||||||||||||| ||| ||| ||| ||| || |
       mRNA:  50 UUCCAAUCCAACCACAAACAAAAAUCAGCCAAU AUG UCC GAC UUC GAG AA 99
    protein:                                     N-Met-Ser-Asp-Phe-Glu-As
                                                    1   2   3   4   5   6

GENOMIC DNA: 351 C AAG AAC CCC AAC AAC GTC CTT GGC GGA CAC AAG GCC ACC CTT CAC AAC C
                 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
       mRNA: 100 C AAG AAC CCC AAC AAC GUC CUU GGC GGA CAC AAG GCC ACC CUU CAC AAC C
    protein:     n-Lys-Asn-Pro-Asn-Asn-Val-Leu-Gly-Gly-His-Lys-Ala-Thr-Leu-His-Asn-P
                    7   8   9   10  11  12  13  14  15  16  17  18  19  20  21  22

GENOMIC DNA: 401 CT AGTATGTATCCTCCTCAGAGCCTCCAGCTTCCGTCCCTCGTCGACATT 450
                 || |
       mRNA: 150 CU A..........................................  152
    protein:     ro-A
                 23

 GENOMIC DNA:451 TCCTTTTTTTTCATATTACATCCATCCAAGTCCCACAATCCATGACTAAC 500

       mRNA:     ..............................................

GENOMIC DNA: 501 CAGAAATATCACAGAT GTT TCC GAG GAA GCC AAG GAG CAC TCC AAG AAG G 550
                              || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
       mRNA: 153 ..............AU GUU UCC GAG GAA GCC AAG GAG CAC UCC AAG AAG G 188
    protein:                  sn-Val-Ser-Glu-Glu-Ala-Lys-Glu-His-Ser-Lys-Lys-V
                              24 25  26  27  28  29  30  31  32  33  34  35

GENOMIC DNA: 551 TG CTT GAA AAC GCC GGC GAG GCC TAC GAT GAG TCT TCT TCG GGC AAG ACC
                 || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
       mRNA: 189 UG CUU GAA AAC GCC GGC GAG GCC UAC GAU GAG UCU UCU UCG GGC AAG ACC
    protein:     al-Leu-Glu-Asn-Ala-Gly-Glu-Ala-Tyr-Asp-Glu-Ser-Ser-Ser-Gly-Lys-Thr-
              36  37  38  39  40  41  42 43  44  45  46  47 48  49  50  51  52
```

```
GENOMIC DNA: 601 ACC ACC GAC GAC GGC GAC AAG AAC CCC GGA AAC GTT GCG GGA GGA TAC AA
                 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
       mRNA: 240 ACC ACC GAC GAC GGC GAC AAG AAC CCC GGA AAC GUU GCG GGA GGA UAC AA
    protein:     Thr-Thr-Asp-Asp-Gly-Asp-Lys-Asn-Pro-Gly-Asn-Val-Ala-Gly-Gly-Tyr-Ly
                  53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69

GENOMIC DNA: 651 G GCC ACC CTC AAC AAC CCC AAA GTG TCC GAC GAG GCC AAG GAG CAC GCC A
                 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
       mRNA: 289 G GCC ACC CUC AAC AAC CCC AAA GUG UCC GAC GAG GCC AAG GAG CAC GCC A
    protein:     s-Ala-Thr-Leu-Asn-Asn-Pro-Lys-Val-Ser-Asp-Glu-Ala-Lys-Glu-His-Ala-L
                    70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85

GENOMIC DNA: 701 AG AAG AAG CTT GAC GGC CTC GAG TAA GCTCAGAGTTCACGAAAGAACCAT 750
                 || ||| ||| ||| ||| ||| ||| ||| ||| ||||||||||||||||||||||
       mRNA: 339 AG AAG AAG CUU GAC GGC CUC GAG UAA GCUCAGAGUUCACGAAAGAACCAU 388
    protein:     ys-Lys-Lys-Leu-Asp-Gly-Leu-Glu-C
                 86  87  88  89  90  91  92  93

GENOMIC DNA: 751 TCGACGAGGGGAAGCACGGGGTTATCTCGTTCGAAACATGGGCCTGGTTA 800
                 ||||||||||||||||||||||||||||||||||||||||||||||||
       mRNA: 389 UCGACGAGGGGAAGCACGGGGUUAUCUCGUUCGAAACAUGGGCCUGGUUA 438

GENOMIC DNA: 801 ATGCAAATGCATAATGGGGAGGATAATGAATCATGAGGTGTACGATATGG 850
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
       mRNA: 439 AUGCAAAUGCAUAAUGGGGAGGAUAAUGAAUCAUGAGGUGUACGAUAUGG 488

GENOMIC DNA: 851 ACGATATTGACGGATCTTAATTTGATGACAGTAATGAAATCACACCATAG 900
                 |||||||||||||||||||||||||
       mRNA: 489 ACGAUAUUGACGGAUCUUAAUUUGAAAAAAAAAAAAAAAAAAAAAAAAA 538
```

Note that the single intron splits in the middle of the codon for amino acid Asn_{24}.

(4.2.5)

a) The sequence TATA is a DNA palindrome; that is, it reads the same (5′ to 3′) on both strands of a DNA duplex. Therefore, it has no inherent direction. This is shown below:

```
5′-CCCCCCCTATATTTTTT-3′
   |||||||||||||||||
3′-GGGGGGGATATAAAAAA-5′
```

If TATA causes transcription to the right of the last A, then this sequence would have to produce two mRNAs: UUUUU... and GGGGG.... This is an unacceptable situation. Since the sequence TATATA is also a DNA palindrome, it would have the same problem. The sequence TATAA is not a DNA palindrome and therefore has an inherent direction.

b) The sequence of the mature mRNA is:

5′ G_{cap}-AAGGCGCUGC**AUG** CAG AAC AUG CCU CUC UUU AAG GUA CUA CUA ACU
AAU GAU GGU **UGA** CGAUACCCUCGGAAUAAAUAAAAAAA 3′

c) The protein sequence is:

H_3N^+-Met-Gln-Asn-Met-Pro-Leu-Phe-Lys-Val-Leu-Leu-Thr-Asn-Asp-Gly-COO^-

d) The 3′ splice junction of intron 1 is lost. The next potential splice site is the CAG (nt 59-61). This changes the 5′ end of the mRNA. Seven nucleotides have been lost between underlined bases; the rest of the mRNA remains the same.

5′: G_{cap}–AAGGCGCUGAAC**AUG** CCU CUC UUU AAG GUA CUA CUA ACU AAU GAU
GGU **UGA** CGAUACCCUCGGAAUAAAUAAAAAAA :3′

This gives a protein sequence (the same sequence, just started at the second AUG):

H_3N^+-Met-Pro-Leu-Phe-Lys-Val-Leu-Leu-Thr-Asn-Asp-Gly-COO^-

e) The 5′ splice junction of intron 1 is lost. Therefore, intron 1 is not spliced out at all. The rest of the mRNA remains the same. This adds a new AUG codon in a different frame, which changes the protein sequence. Added nucleotides are underlined.

5′: G_{cap}–AAGGC GCUGG U**AUG**,U CC,GAA **UAG**CA UGCAG AACAU GCCUC etc.

The resulting protein sequence: H_3N^+-Met-Ser-Glu-COO^-

f) The 3′ splice site in intron 2 is lost. Therefore, the splicing machinery will search for the next CAG or UAG it can find. The next CAG is at the 3′ end of intron 3. This results in a deletion of all the sequences of exon 2. The deletion occurs between the underlined nucleotides:

5′: G_{cap}–AAGGC GCUGC **AUG,**CA G,AAC,A UG,CCU CUU,AA C,**UAA**U GAUGG
UUGAC GAUAC CCUCG GAAUA AAUAA AAAAA :3′

The resulting protein sequence (a truncated protein):

H_3N^+-Met-Gln-Asn-Met-Pro-Leu-Asn-COO^-

g) This creates a new 3′ splice site within the intron. This adds five nucleotides to the mRNA (underlined):

5′: G_{cap}–AAGGC GCUGC **AUG,**CA G,AAC,A UG,CCU CUC,UU U,AAG,G UA,CUA
CUA,CA G,**UAA**C UAAUG AUGGU UGACG AUACC CUCGG AAUAA AUAAA
AAAA :3′

This causes premature termination of the protein chain:

H_3N^+-Met-Gln-Asn-Met-Pro-Leu-Phe-Lys-Val-Leu-Leu-Gln-COO^-

h) This mutation destroys the promoter TATAA sequence. Therefore, there will be no transcription of the gene. No mRNA will be made, and no protein will be made.

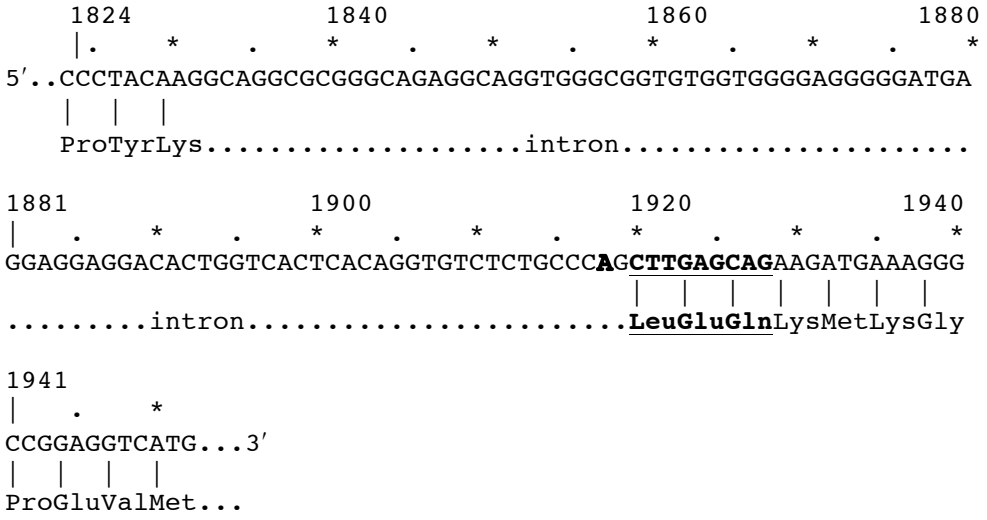
(4.2.6)
a) If you start with the first three nucleotides, CCC, they encode proline, then tyrosine (TAC), and lysine (AAG). However, the next three nucleotides, GCA, do not encode lysine. The next lysine codon (AAA or AAG) starts at position 1929. After that point, the coding sequence is uninterrupted for the region of the protein shown. This is shown below (coding sequences shown **bold and underlined**). The noncoding sequence must be the intron, which begins at position 1833 and ends with position 1928.

```
    1824            1840                1860            1880
    |.    *    .    *    .    *    .    *    .    *    .    *
5′..CCCTACAAGGCAGGCGCGGGCAGAGGCAGGTGGGCGGTGTGGTGGGGAGGGGGATGA
    |  |  |
    ProTyrLys.....................intron.....................

1881               1900                1920            1940
|   .    *    .    *    .    *    .    *    .    *    .    *
GGAGGAGGACACTGGTCACTCACAGGTGTCTCTGCCCGGCTTGAGCAGAAGATGAAAGGG
                                                    |  |  |  |
.........intron................................LysMetLysGly

1941
|   .    *
CCGGAGGTCATG...3′
|  |  |  |
ProGluValMet...
```

b) The intron starts and ends with (5′) GC.....intron.....AG (3′). Even though this does not follow the "rule," the protein sequence data are more compelling, since it bears directly on where the splice sites are located. The "rule" is only a consensus based on many sequences; there are always variations.

c) i) Since a single base-pair substitution caused three amino acids to be inserted in the region from which the intron was removed, it is logical to suppose that some alteration of the splicing signals has caused less of the intron to be removed than would be removed in wild type. Therefore, the start of the intron (5′ splice site) must have moved to the right or the end of the intron (3′ splice site) has moved to the left. Looking at the sequence, the intron 3′ splice site must have moved nine nucleotides to the left to encode the amino acids given in the problem. This is shown below (the new coding region is shown **bold and underlined**). The first and last nucleotides of DNA that encode the mutant intron are 1833 and 1919, respectively.

ii) To make the 3′ splice site be at position 1919, nucleotide 1918 must have changed from a G to an A, making a new 3′ splice site. This base is shown in outline type on the next page.



Computer Activity 2: Gene Explorer (GeneX)
1) There are no solutions for this part.

2) a) Any of the following bases: 20-26, 55-67, or 92-101. These are in the "coding region"; you could also have called this part the "exon," but that term is less precise.

b) This is not possible. Except for the poly(A) tail, all the bases in the mature mRNA must have counterparts in the pre-mRNA. In the normal gene, the poly(A) tail is not translated.

c) Any of the following bases: 27-54 or 68-91. These are in the introns.

d) Any of the following bases: 11-19 or 102-114. These are the untranslated regions of the exon.

3) a) They are not encoded in the DNA; they are added by an enzyme called poly(A) polymerase without using base-pairing. This is an interesting exception to the DNA/RNA rules: these A's are added to the 5′ end of the mRNA in a 5′ to 3′ direction; they are also added without base-pairing–there are no T's in the DNA that they pair with.

b) Because the splice junctions are seamless–there is no sign that the intron was there–so the ribosomes read right over where the splice junction was.

c) Intron 1 is 28 nucleotides long (not a multiple of 3); Intron 2 is 24 (a multiple of 3). Since introns are spliced out before being read by the ribosome, the length of the intron does not need to be a multiple of three nucleotides.

Part III

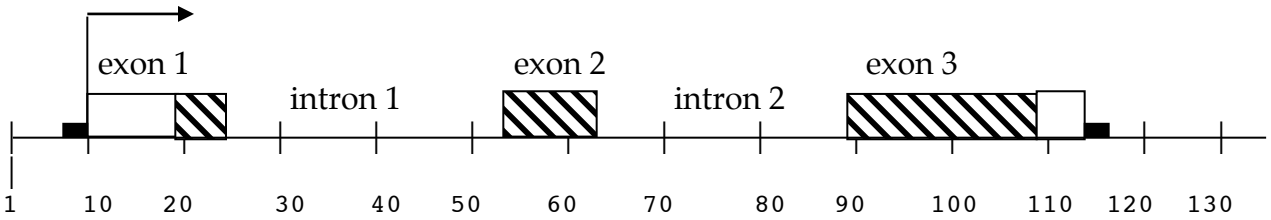
4) and 5) You can make a single-base insertion, deletion, or substitution without changing the protein sequence in the following locations:

- Before the promoter: bases 0-5. This is not part of the gene.
- After the terminator: bases 120-127. This is also not part of the gene.
- In the introns, as long as your mutation doesn't alter the splice sites at the start and end of the intron or create a new splice site: bases 32-49 and 73-86. These do not code for the protein since they are not part of the mature mRNA.
- In the untranslated parts of the exons, before the start codon or after the stop codon, as long as they don't make a new start codon: bases 11-19 and 105-114. These do not code for protein.

You cannot make insertion or deletion mutations in the coding region without changing the protein sequence. Substitution mutations are allowed at certain sites in the coding region if they don't change the amino acid (silent mutations): for example, changing base 101 from T to A.

8) N-Met-Pro-**Cys-Arg-Met-Ser-Ser-Glu-Asp-Leu-Lys-Lys-Val-C**
Everything after the Pro at position 2 is altered.
This is a FRAMESHIFT mutation.

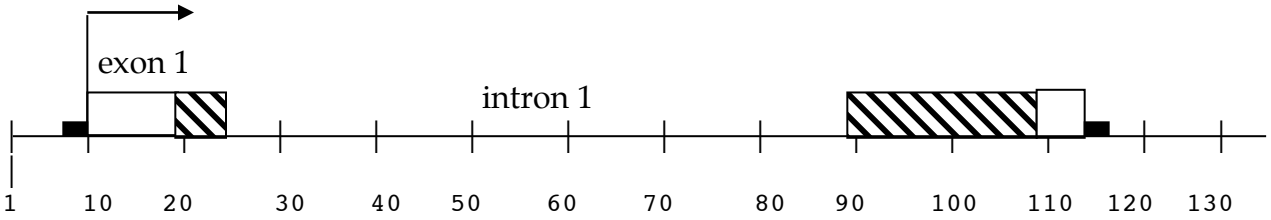
9) This is the same map as the normal gene with only the position of the stop codon changed.



10) The deleted base changed the reading frame.

13) N-Met-Pro-**Cys-Glu-Asp-Leu-Lys-Lys-Val-C**
Everything after the Pro at position 2 is altered.
This type of mutation does not have a name.

14) Note that this map is very different from the normal gene.



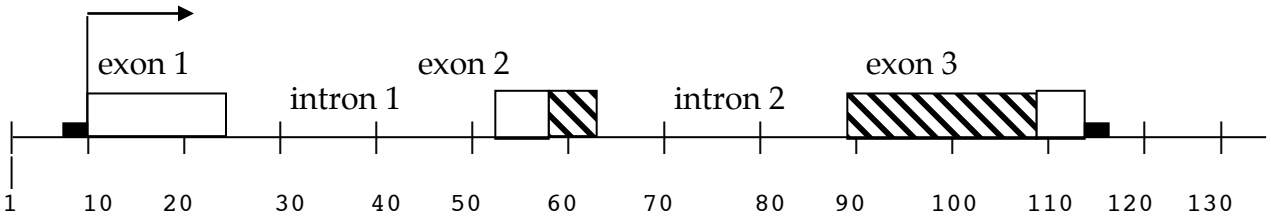
15) The mutation destroyed the "end of intron 1" (a.k.a. "3′ splice site") sequence so the splicing machinery went looking for the next downstream "end of intron" sequence. It therefore skipped exon 2 entirely!

18) **N-Met-Ser-Ser-Glu-Asp-Leu-Lys-Lys-Val-C**

Everything is altered.

This type of mutation does not have a name.

19) (same map as before with just the position of start and stop changed)



20) This mutation destroyed the start codon so the ribosome went looking for the next AUG. This makes a totally new protein with a different reading frame and a different stop codon. Note that exon 1 is completely nontranslated.

(22) and (23) To have this effect, the mutation must change a regular amino acid codon to a stop codon. There are several such mutations, for example, changing 55 from T to A; changing 58 from C to A; and changing 96 from G to T.

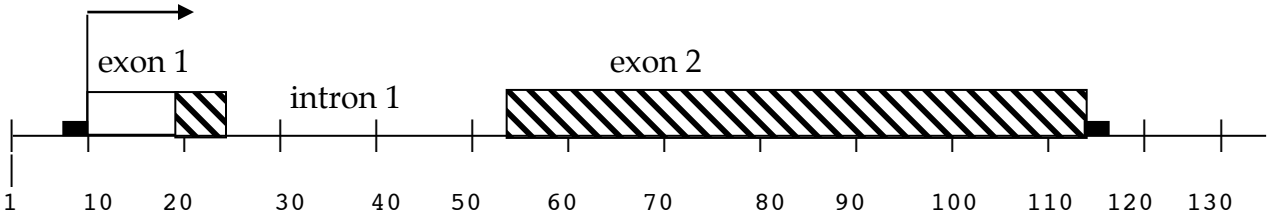
(25) and (26) To have this effect, the mutation must be in the "start intron" signal at the 5′ end of an intron. That is:

- Changing bases 27-31. These eliminate the "start intron" signal for intron 1; this gives a mature mRNA about 92 nt long because intron 1 is not spliced out.
- Changing bases 68-72. These eliminate the "start intron" signal for intron 2; this gives a mature mRNA about 87 nt long because intron 2 is not spliced out.

Note that altering the "stop intron" signal at the 3′ end of an intron (50-54 and 87-91) causes the following exon to be skipped so the mutant mRNA is shorter than normal.

(28) and (29) Only a mutation in the promoter or terminator will completely abolish mRNA and protein synthesis from the gene.

31) Delete base 68. This knocks out the start of intron 2 so it is not spliced out. This adds many more amino acids to the protein. The resulting gene is:



33) For each of these, there are several possibilities; we will describe only one.

a) Delete 26. This causes a frameshift. The resulting reading frame does not stop until the last codon in exon 3 and produces a protein that is three amino acids longer than normal.

b) Click on 60 and insert a T. This produces a stop codon that results in a shorter protein.

c) Delete 52. This inactivates the "end intron 1" sequence. This causes the splicing machinery to skip over exon 2 to the next "end intron" sequence.

(5) CHALLENGE PROBLEMS

(5.1)
a) With a circular chromosome, the DNA is continuous–it has no end. This means that there will always be DNA from which to make the RNA primer for the lagging strand.

b) A small piece of the telomeric DNA is lost during replication, so telomeres do not solve the problem of shrinking chromosomes but lessen the impact. The loss of telomeric DNA does not cause harm to the organism, because the telomeric DNA does not encode any genes. Its function is to protect the rest of the chromosome from being slowly lost (from the ends inward) during successive rounds of replication. The shrinkage of the telomere is later compensated by the action of telomerase.

(5.2)
a) Three key differences between DNA and RNA are:

1) DNA is usually double stranded whereas RNA is mainly single stranded.

2) DNA contains thymine and RNA contains uracil.

3) The ribose component in DNA lacks a 2′-OH group; however, this hydroxyl group is present in the ribose moiety found in RNA.

b) Hydrogen bonds allow base pairing to occur between the two strands of DNA in a double-stranded DNA helix.

c) You would expect sequence (ii) to denature at a higher temperature. It contains a higher number of G/C base pairs. Each G/C base pair is held together by three hydrogen bonds, as opposed to two hydrogen bonds for A/T base pairs.

(i)
```
ATAGTATTC
|||||||||
TATCATAAG
```

(ii)
```
GACGCGGTG
|||||||||
CTGCGCCAC
```

d) The hydrophobic portions of each base interact with the hydrophobic regions of bases stacking above and below in the helix. These van der Waals and hydrophobic interactions contribute significantly to the structural stability of a double-stranded DNA molecule.

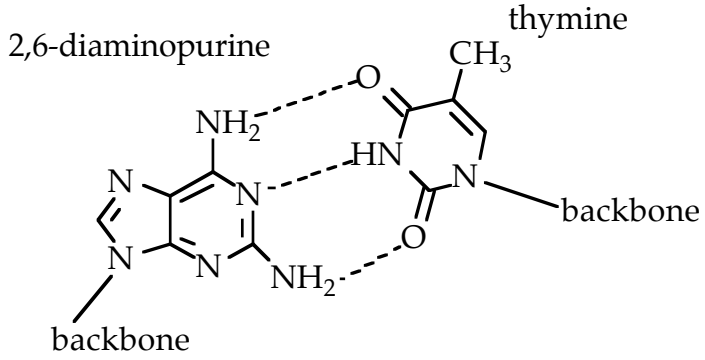
e) The A and G bases are purines that have two aromatic rings and are larger than the pyrimidines. The T and C bases are pyrimidines that have one aromatic ring. The distance between the backbones of each strand of the double helix accommodates a purine hydrogen bonded to a pyrimidine. A base-pair mismatch of A-G would increase the distance between the backbones, and a base-pair mismatch of C-T would decrease the distance between the backbones.

f) In a cell that is at physiological pH, the overall charge of a double-stranded DNA molecule is highly negative because of the charged phosphate groups on the DNA backbone. (The pKa of the phosphate group is ~1.0; hence, the reason why the molecule is called deoxyribonucleic acid. In a cell the negative charges are neutralized by NaCl.)

g) DNA replication should halt or be slowed down by the addition of the AZT nucleotide. Since the AZT nucleotide lacks a 3′-OH group, every time it is incorporated into an elongating DNA strand it stops DNA replication after that point.

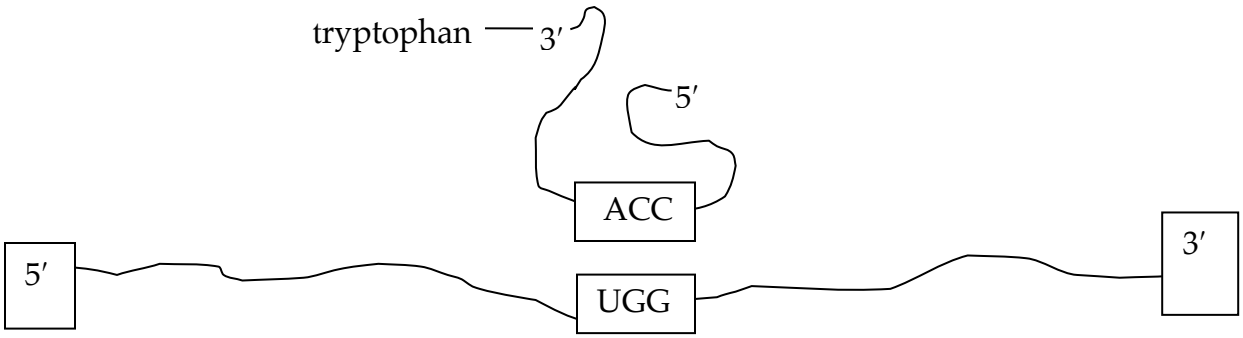
If AZT is incorporated into the DNA of cells that are undergoing cell division, it would inhibit DNA replication, thus leading to cell death. Cancer cells divide rapidly and thus would be most highly affected by AZT. The HIV nucleic acid polymerase preferentially incorporates AZT into its genome, thus inhibiting the replication of the viral DNA.

h) The base 2,6-diaminopurine would most likely form a base pair with thymine.



(5.3)

a)



b) The lower strand is used as the template strand. When transcribed, the sequence for the anticodon in the tRNA should be 5′ CCA 3′. Therefore, the lower strand must be used as the template:

3′ …GGT… 5′ used as template
5′ …CCA… 3′ tRNA

Integration Solutions

(1)

a) With no active glycosyltransferase of either type, an ii individual would not be able to add any sugars to the O form of the lipopolysaccharide. Thus, the only lipopolysaccharide present would be the O type.

b) The enzyme encoded by the i allele is essentially inactive. Having one copy of the enzyme (encoded by the I^A or I^B allele) that adds sugars onto the O antigen must be sufficient to produce enough of the A or B antigen on the surface of red cells to provoke an immune response.

c) Here, both enzymes encoded by the I^A and I^B alleles are active, and both reaction 1 and reaction 2 will occur simultaneously. Thus, the cells will be producing both A and B antigen, and the red cells will have both molecules on their surfaces. Consequently, both antigens will be available for immune response.

d)

```
                          .Leu Val Val Thr Pro Trp...
Sequence of I^A allele:   ...C GTG GTG ACC CCT T...

                          .Leu Val Val Pro Leu...
Sequence of i allele:     ...C GTG GTA CCC CTT....
```

e) Note that the reading frame must be the same in both sequences, starting from the left.

```
                                266 267 268
                          ..Tyr Tyr Leu Gly Gly Phe Phe...
Sequence of I^A allele:   ...AC TAC CTG GGG GGG TTC TT...

                                266 267 268
                          ..Tyr Tyr Met Gly Ala Phe Phe...
Sequence of I^B allele:   ...AC TAC ATG GGG GCG TTC TT...
```

f) The difference between the type-A and type-B sugars is mostly size; both groups are capable of forming hydrogen bonds with the enzyme. The type-A sugar is larger than the type-B sugar. Thus, the side chains in the binding site of the type-A enzyme are smaller (Leu and Gly) than the corresponding side chains in the type-B enzyme (Ala and Met). Thus, the larger substrate can only fit into the larger binding site while the smaller can fit into the smaller binding site.

g)

i) This would change one of the amino acids at the binding site from very small (Ala or Gly) to very large (Arg). This would make the binding site too small to accommodate the substrate, resulting in an inactive enzyme.

ii) This would result in a substantially shorter glycosyltransferase enzyme which would likely be missing critical parts and be inactive.

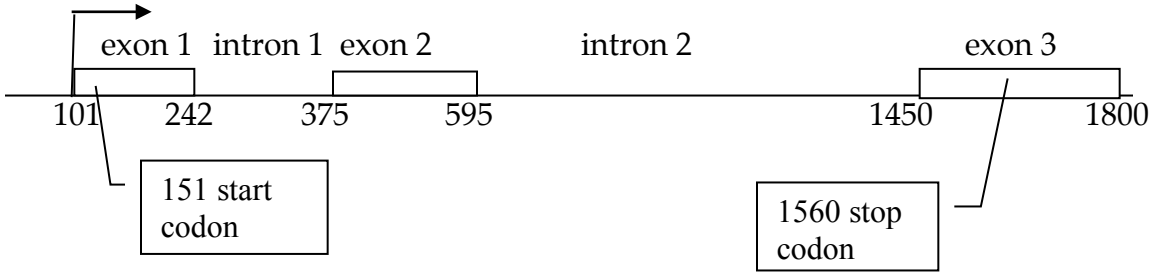
h)

i) The differences in sequence must eliminate the UDP-X substrate specificity of the substrate-binding pocket, so that it now binds either UDP-Gal or UDP-GalNAc.

ii) If you had a patient with type AB who had parents of types AB and O, this would indicate the presence of a *cis*-AB allele. Without the *cis*-AB allele, this cross would be: $I^AI^B \times ii$, which cannot give an AB child. The only explanation (barring parental infidelity) is $I^{cisAB} \times ii$, giving a I^{cisAB} i child with type AB blood.

(2)

a) Your map should look roughly like this:



b)

A1: UGG ⇒ UAG nonsense mutation.
This would result in a protein only ~15 amino acids long. The vast majority of the protein would be missing, so it would be nonfunctional.

A2: AAG ⇒ UAG nonsense mutation.
This would result in a protein only ~16 amino acids long. The vast majority of the protein would be missing, so it would be nonfunctional.

A3: CAG ⇒ UAG nonsense mutation. This would result in a protein only ~35 amino acids long. The vast majority of the protein would be missing, so it would be nonfunctional.

A4: Frameshift. Hint: Use the translation below the DNA instead of the codon table. The protein is now N-Met-Val-His-Leu-Thr-Pro-GLY-Arg-Ser-Leu; the reading frame is shifted to frame b starting with the Arg (the GLY comes right at the deletion). The remaining amino acids (until the early stop at about 205) will all be different than in the normal protein. This protein will therefore be nonfunctional.

A5: Frameshift to frame c. N...-Glu-Glu-Val-Cys-.... (frame c starting from Leu). The amino acids will all be different from those in the normal protein up to a premature termination at 217 or so. This protein will therefore be nonfunctional.

A6: Frameshift to frame c. N...-Glu-Glu-LYS-Val-Cys.... (frame c starting from Leu). The amino acids will all be different than those in the normal protein up to a premature termination at 217 or so. This protein will therefore be nonfunctional.

Each of these mutations will result in a recessive phenotype because in each case the mutant gene does not produce functional protein. One functional copy of the β-globin gene produces enough protein to make healthy red blood cells. Thus, the heterozygote is normal.

c)
B1: GGC (Gly, small and hydrophobic) ⇒ CGC (Arg, big and hydrophilic). The side chain is buried in the hydrophobic core, an unfriendly place for a hydrophilic amino acid. This substitution disrupts the β-globin structure and makes it inactive.

B2: GCC (Ala, small and hydrophobic) ⇒ GAC (Asp, small and hydrophilic). The side chain is buried in the hydrophobic core, an unfriendly place for a hydrophilic amino acid. This substitution disrupts the β-globin structure and makes it inactive.

B3: AAG (Lys, big, hydrophilic, and positively charged) ⇒ GAG (Glu, big, hydrophilic, and negatively charged). If the side chain of this amino acid was involved in an interaction with another, this substitution might be expected to disrupt the function of the protein. Normal function is observed, so the side chain is likely on the surface due to its hydrophilicity, but the charge may not be important.

B4: GUC (Leu, medium-sized and hydrophobic) ⇒ UUC (Phe, big and hydrophobic). The side chain is in the interior of the protein and the larger size likely alters the folding of the protein.

B5: CAU (His, medium and hydrophilic) ⇒ CGU (Arg, big and hydrophilic). The His side chain binds to the heme iron, so size alters the binding.

B6: AAA (Lys, big, hydrophilic, and positively charged) ⇒ GAA (Glu, big, hydrophilic, and negatively charged). The side chain is near COO^- of the heme. Changing the amino acid to a negatively charged one will repel heme. Without heme, the structure is really disrupted and the proteins stick together.

Each of these mutations will result in a dominant phenotype because one mutant copy of the gene produces enough mutant protein to aggregate even in the presence of normal protein.

(3)
a) Four of the DNA codons for arginine (CGA, CGG, CGC, CGT) contain the highly mutable CG sequence. If the CG in the CGT or CGC sequences were mutated to TG as described, they would become TGT or TGC, which code for cysteine. Thus, the first factor would make these mutations likely. Also, changing the highly hydrophilic

arginine to the hydrophobic cysteine would likely have a big effect on protein structure. Therefore, individuals with such a mutation would likely show a phenotype.

b) This is similar to (a). The required mutation, CGG to TGG, is likely. Also, changing the highly hydrophilic Arg (hydrophilic) to Trp (hydrophobic) would have a big effect on protein structure.

c) The required mutation, CGT or CGC to CAT or CAC, is very likely, but both Arg and His have hydrophilic and positively charged side chains. It is not surprising that such a mutation occurs, but more that the resulting amino acid change has a noticeable effect on protein structure.

d) The required mutation GTX (where X is A, G, C, or T) to CCX requires two base changes, neither of which involves the highly mutable CG sequence. Thus, the mutation is unlikely. Also, the mutation, from one hydrophobic amino acid to another, is unlikely to have a dramatic effect on protein structure.

e) The required mutation (CTT, CTC, or CTA to ATT, ATC, or ATA) requires only one base change, so it is reasonably likely. However, both amino acids have very similar side chains, so this mutation would not be expected to have a large effect on a protein's structure. It is surprising that individuals with such a mutation show a phenotype at all.

f) The required mutation (GGX to UUX) requires two different mutations, neither of which involves a CG sequence. Therefore, this would not be expected to be very frequent. Also, the mutation makes a dramatic change in the amino acid. Replacing the amino acid with the smallest side chain with an amino acid with a large side chain would have a big effect on protein structure.

(4)

a)

5′-	G	A	A	U	C	G	C	U	A	C	A	A	- 3′

b)

```
5′ TGCCA TCCGA TTGGT GTTCC TTCCA TGAAG GATGC ACAAC GCAAA 3′
3′ ACGGT AGGCT AACCA CAAGG AAGGT ACTTC CTACG TGTTG CGTTT 5′

5′ TACAC GCTTA GCTGA CTATA AGGAC GAATC GCTAC AACGA TGCGA 3′
3′ ATGTG CGAAT CGACT GATAT TCCTG CTTAG CGATG TTGCT ACGCT 5′

5′ TGCCA TCCGA TTGGT GTTCC TTCCA TGAAG GATGC ACAAC GCAAA 3′
3′ ACGGT AGGCT AACCA CAAGG AAGGT ACTTC CTACG TGTTG CGTTT 5′
```

c)

d) You must add ribosomes and aminoacyl-tRNA molecules (charged tRNA molecules).

e) Puromycin mimics a charged tRNA. It would not directly affect transcription.

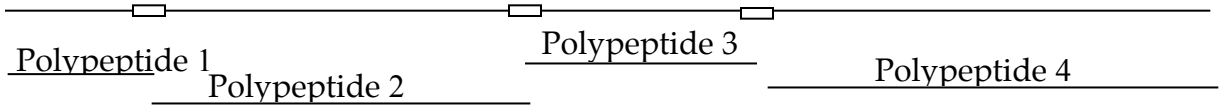
f) Puromycin mimics a charged tRNA. When incorporated into the growing polypeptide, it interrupts translation and results in truncated proteins.

g)

i) In test tube 1 (no puromycin) you get a polypeptide that is 100 amino acids long. Each amino acid represents a codon on the mRNA. A codon is three nucleotides long, so the mRNA was at least 300 nucleotides.

ii) If puromycin does **NOT** recognize a specific codon and is in limiting amounts, then one expects that some translation events will not be interrupted and those that are will be interrupted in a random fashion. The population of polypeptides in tube 2 will include polypeptides of all sizes, i.e., dipeptides, tripeptides, etc.

iii) If puromycin recognizes a specific codon that occurs three times in the mRNA and these codons are not evenly spaced, then you might expect to find four types of polypeptides that are each different lengths.



h)

i)

ii) The alternative sequence has an insertion.

iii) You would not expect this DNA sequence to encode a protein that binds serotonin because all amino acids after the insertion are altered. In this case three of the five amino acids involved in the binding, Ser, Val, and Glu, are after the site of the insertion.

i)

i)

```
5′ ACC AAT GGA CCA GCA GGA AGC GGG GTA GCT GAT TAC 3′
3′ TGG TTA CCT GGT CGT CCT TCG CCC CAT CGA CTA ATG 5′
```

ii) This sequence has a substitution mutation.

iii) You would expect this DNA sequence to encode a protein that binds serotonin. The substitution replaces a Glu with Asp. These two amino acids are very similar.