# Biol 213 Genetics (13 September 2000)
# Relationship between Genes and Proteins

Wednesday marks the changing of the guard: Brad marches out, and I march in. You might view this transition warily, as an obstacle to overcome. After all, you've just gotten used to the way Brad does things, and here comes a new way.

It's true, Brad and I approach things somewhat differently. This not because one way is better than the other but rather because we each have different strengths and we draw on them in different ways to provide the best possible learning experience we can. You may be sure, however, that while the surface may appear different at times, underneath we share the same goals for the course and are constantly consulting with one another as to the best way to achieve them. Rather than an nuisance, you might view this as an opportunity to skim the cream off two different approaches.

Let me provide what help I can to help you anticipate the change. I am quieter than Brad. I draw on your energy rather than try to fill the room with my own. I will ask more questions, try more to solicit a conversation typical of smaller classes. I'll do so with fewer tools at my disposal, since my capacity for names is much smaller than Brad's, despite my best efforts. I hope you'll make allowances.

My approach will work best if you come prepared to answer the study questions you can answer and to identify problems in answering the questions that stump you. It would help also if you recognize the limitations of the ridiculous room we find ourselves in and speak up – not just to me but to your colleagues as well.

One more thing. Brad is not marching off to Virginia Beach, just to the back of the room. He remains an active participant and will continue attend class, just as I did while he stood up front (those of you in the 11:30 section may not have realized this since I went to the earlier section). He and I are both available outside of class, prepared to discuss ANY aspect of the class.

**OUTLINE**

**I. DNA**
- A. DNA Replication (continued) (pp.295-297)
- B. Plasmids and viruses

**II. Relationship between genes and protein**
- A. The issue (pp.312-316 skim)
- B. Archibald Garrod and inborn errors of metabolism (p.316)
- C. Beadle and Tatum's experiments (pp.316-321)

**III. RNA and an overview of gene expression** (pp.238-240; 321-325)

*A good strategy might be to do the suggested reading, making an outline as you go, and then go through the notes below, where I read the chapter along with you.*

**I. DNA**

I.A. DNA Replication (continued) (pp.295-297)

Evolution places a premium on what works, not on what is elegant. As a result, cells are full of clever fixes, using the materials and capabilities at hand. Very often the solution works in ways that seem to us circuitous, not at all how we would do it if we were designing things. Evolution rewards creatures that win – ugly wins are just fine.

So it is with DNA replication. Take a look at Figure 11-6. It would be so simple if one DNA polymerase made one strand 5' to 3' and another DNA polymerase made the other 3' to 5', but that would require the invention of two wholly different enzymes. The solution that was found some billions of years ago exploited one enzyme, used in a complicated way.

**SQ1. Identify the DNA polymerases in the figure. Which way are they polymerizing? (answer in words but also by pointing).**

**SQ2. The brown strand represents parental DNA. What do the blue and red strands represent? Why do the two leftmost blue arrows have red attached to them but the next arrow to the right does not?**

**SQ3. What does the light blue box represent? What is it for? What would happen to the picture if it didn't exist?**

**SQ4. What does the blue circle represent? What is it for? What would happen to the picture if it didn't exist?**

**SQ5. What do the orange circles represent? What are they for? What would happen to the picture if they did not exist?**

**SQ6. What does the blue oval represent? What is it for? What would happen to the picture if it did not exist?**

**SQ7. Describe how the picture will look a few moments in the future.**

**SQ8. Describe how the picture would change if cells did use a 3' to 5' DNA polymerase.**

From the looks of Figure 11-6, the leading strand is made continuously but the lagging strand is made haphazardly. There seems always to be the risk that the leading strand will be made a segment of the lagging strand will be forgotten. This is not the case. In fact, both strands are made simultaneously by the same enzyme complex. Figure 11-12 shows how this occurs. I have great difficulty making sense out of this figure. The important thing to see is that the two parts of the DNA polymerase complex (the purple blobs labeled DNA pol III), progress in the same physical direction, but because of the kink in the DNA, one strand is made in the direction of overall replication (towards helicase) while the other is made away from the direction of overall replication (away from helicase).

**SQ9. Do you see how this works?**

I.B. Plasmids and viruses

*Big dogs have little fleas to torment and bitum,*
*And little fleas have littler fleas, on ad infinitum*

I'm not sure I got this poem right, and I forget who wrote it, but it works as a description of life. The bacteria that live on and in us -- some for good, some for ill -- have their own parasites they suffer. Figure 1 gives you an idea of the scales involved. Bacteria have roughly a thousand-fold less DNA than humans and most higher eukaryotes. Their plasmids and viruses (also called bacteriophage or simply phage) have roughly ten- to a hundred-fold less DNA than that. We use the relatively small size of plasmids and viruses to our advantage. It is much easier to manipulate their DNA than the huge amount of DNA in cellular organisms.

I used the term "cellular organism" to distinguish prokaryotes and eukaryotes from plasmids and viruses, leaving it to philosophers to decide whether or not they too constitute forms of life. If they do, they can be considered cousins. Many viruses are like plasmids in being able to live quietly within their host, often conferring selective advantage. Both are adapted to the promiscuous life, switching hosts as the opportunity arises. Sometimes, bacterial DNA gets caught up in the transfer of DNA, and in this way host DNA can move to different bacteria, even those very distant on the evolutionary scale. When plasmids are involved, the transfer is called conjugation. When viruses are involved, it's called transduction. Bacteria may not have sex, but they do exchange DNA.

We think of viral infections as an unmitigated evil. Yet, if they are used to transfer DNA from cell to cell or even organism to organism, what is detrimental to most individuals may be of benefit to the species as a whole. Who knows what useful

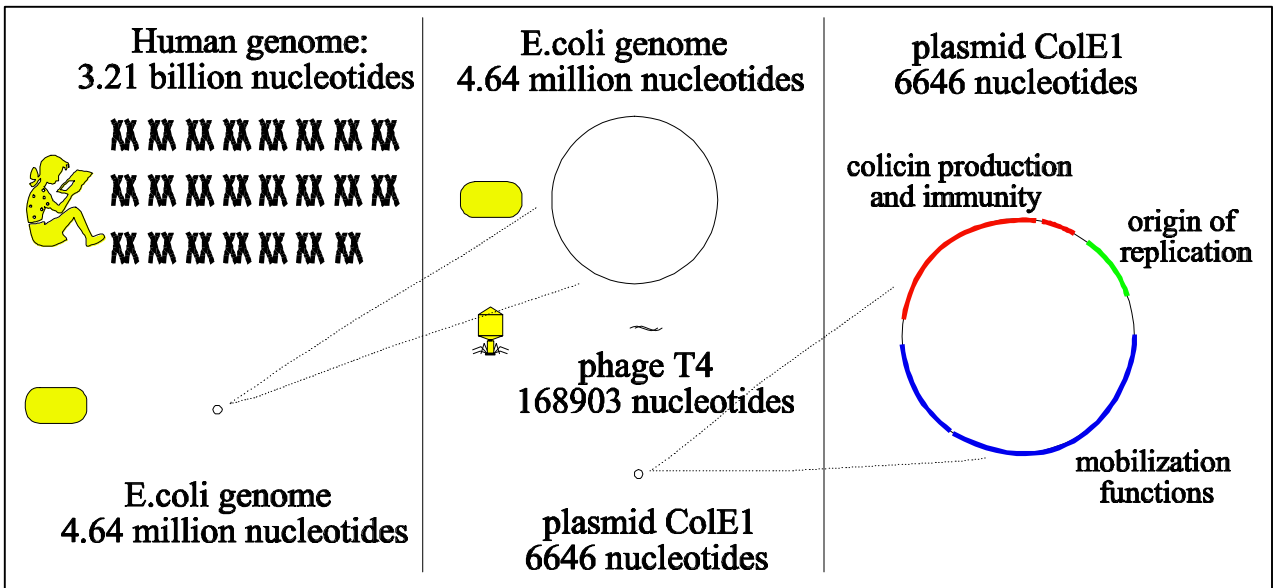


**Figure 1: Comparison of genome sizes, from humans to plasmid.** The genomes are not drawn to scale (if they were, the *E.coli* genome on the left would be a pinprick). The genes of plasmid ColE1 are grouped by function. Mobilization genes are those necessary to move the plasmid from one strain to another, through the conjugal apparatus encoded by another plasmid.

genes you might pick up in the molecular flea market?

Natural plasmids and viruses carry the genes required for their replication and either infection or conjugal transfer. In addition, they often carry genes that make them valuable to their hosts. Viruses are able to protect their hosts from infection of some other viruses. Plasmids carry a wide variety of nonessential genes. Plasmid ColE1 (shown in Figure 1), for example, encodes the secreted protein colicin E, which kills some bacteria around the host, thus eliminating the competition.

We have disemboweled natural plasmids, throwing out all but the genes necessary for replication, to create man-made plasmids useful in propagating genes of our choosing. The plasmid pUR3 (described in your lab manual) is an example of a man-made plasmid. The *lac* genes from *E. coli* carried by pUR3 were put into a ColE1-like plasmid. The 6300 nucleotides of the genes can be pretty hard to find in the 4.6 million nucleotide genome, but it's a major chunk of pUR3. This makes it easy to isolate large quantitities of the genes, facilitating the manipulation of the genes or further dissection, as you do in Lab 2.

## II. Relationship between genes and proteins

II.A. <u>The issue</u>

Add up the first two weeks of the semester and you'll see that we have established that:

- Enzymes (protein) are the active principles that determine the form and function of a cell
- Genes (DNA) are the medium of inheritance

Since clearly the form and function of a cell is inherited, it follows that somehow DNA must determine proteins. How?

Not a fair question. From what you've learned about the structure of DNA and protein and from what you've picked up at various points in your travels, you already have a pretty good idea of the answer. To those who grappled with the question in the early part of the century, the answer was far from clear. Here are some possible alternatives:

1. Genes determine shapeless protoplasm, that wrap itself around chemicals from the environment to form active sites. Call this the one-gene-many-enzyme hypothesis.
2. Genes determine the components of proteins, like the nuts and bolts that make up an automobile. They are assembled differently to make different protein. Call this the many-genes-many-enzymes hypothesis.
3. A gene's sole task is to specify the structure of a single enzyme. Call this the one-gene-one-enzyme hypothesis.

II.B. Archibald Garrod and inborn errors of metabolism (p.316)

The text relates Archibald Garrod's insights into the nature of alkaptonuria, a disease caused (we now know) by a mutation in the gene encoding homogentisic acid oxidase. The text describes the disease as the accumulation of homogentisic acid in the urine. The compound gets oxidized and forms a black compound. Black pee is actually the least of the person's worries – alkaptonuria is also associated with arthritis and obstruction of the urinary tract – but it is a unique symptom that makes the disease easy to spot.

**SQ10. Why does homogentisic acid accumulate?**

The key insight Garrod made was that the disease appeared to be inherited as a recessive trait. Garrod thus connected inheritance to a defect in a specific enzyme.

**SQ11. Suppose that the inferred mutation renders homogentisic acid oxidase totally inactive (e.g., the mutation is in the active site). How much enzyme activity would you expect from a person who is homozygous for the mutation?**

**SQ12. What would you expect to be the phenotype of a person who is heterozygous for the mutation? Why?**

**SQ13. Consider each of the three hypotheses listed above. Try to square Garrod's findings with each of them.**

**SQ14. By analogy with alkaptonuria, what phenotype would you expect from PKU? See Figure 12-2.**

Phenylketonuria (PKU) is a common inherited defect resulting from the loss of phenylalanine hydroxylase. Phenylalanine accumulates to a point where a little used biochemical pathway, taking phenylalanine to phenylpyruvate (phenylketone). This compound is toxic to developing nerves, leading to mental retardation.

**SQ15. If a person with PKU can be identified shortly after birth, there is an effective treatment: a special diet with low levels of phenylalanine. Why does this diet prevent the effects of PKU?**

II.C. Beadle and Tatum's experiments (pp.316-321)

The experiment described in the text might at first seem perfectly straightforward, if you don't look at the details, or next to impossible, if you do. So let's look at the details. The choice of *Neurospora* as the experimental organism was crucial. Suppose instead that they chose microscopic humans. What's the chance they would find mutants that were defective in a particular enzyme? Consider that mutations in any specific protein is very rare. The vast majority of the cells irradiated by Beadle and Tatum escaped without any detectable mutation. Consider too that we, like almost every other organism you look at in daily life, have two copies of the gene that specifies a specific protein. Hitting that gene twice by irradiation would be very unlikely.

Beadle and Tatum instead irradiated *Neurospora* grown in its haploid form. They could have done this with microscopic humans too, irradiating submicroscopic sperm or eggs. But they could not have done the next step: growing up the irradiated haploid form to examine their phenotypes. Our own haploid cells are incapable of independent growth. (Imagine the effect on fertilization clinics if you could grow cultures of sperm and eggs and test them for phenotype prior to fertilization).

**SQ16. Suppose that haploid *Neurospora* also were incapable of growth. What would Beadle and Tatum have had to do to find mutants?**

Testing haploid cultures for mutant phenotypes is so much easier than testing diploid cultures formed by combining mutant haploid cells with wild-type haploid cells. All Beadle and Tatum had to do is to grow the irradiated haploid spores in minimal medium. The rare spores unable to grow in minimal medium were the mutants they sought.

**SQ17. What does "minimal medium" mean? How does it differ from "complete medium"? How does it differ from "water"?**

What does this experiment say regarding the relationship between gene and protein? Any experiment of value must be able to yield more than one possible answer.

**SQ18. What would have been the results if the one-gene-many-enzyme hypothesis were true?**

**SQ19. What would have been the results if the many-gene-many-enzyme hypothesis were true?**

Beadle and Tatum's conclusions relied on their ability to say that their mutants had one and only one mutation. How did they know this? The text says, "*This 4:4 ratio indicates that a single mutation is causing a vitamin requirement*" (p.318, step 5, right column). The ratio is informative because of the peculiar way that *Neurospora* propagates and is another reason why the choice of organisms was critical. How do they propagate? Don't ascus. (Sorry). Clearly the clue to the 4:4 ratio lies in the middle of p.318, step 4, where you see eight spores lined up in what is called an ascus. How would you find out what an ascus is all about? A trip to the index of the text directs you to p.123, and the middle paragraph points to Figure 5-12b, where you see how a heterozygotic fungus would produce spores: half of the spores would carry the wild-type allele and half would carry the mutant allele. Eight spores, half wild-type, half-mutant. . . that's 4:4.

**SQ20. It is evident that the spores shown within the ascus in step 4 arose from a heterozygous diploid *Neurospora*. How did Beadle and Tatum make the heterozygous strains?**

Now we can appreciate the importance of the manipulations described in step 3. Crossing the irradiated haploid spores formed in step 2 with a wild-type haploid strain produces the desired heterozygotes.

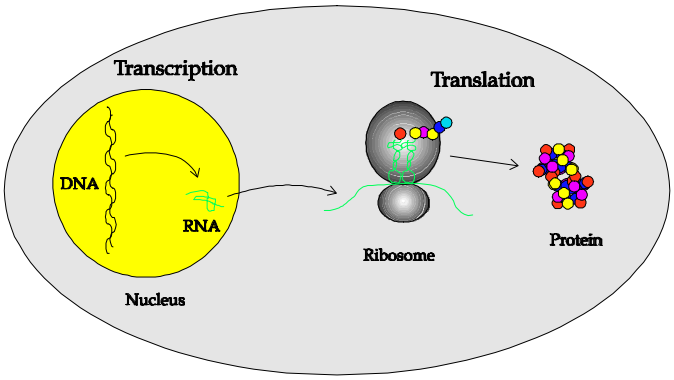


**Figure 2. Information flow through a eukaryotic cell.** DNA is transcribed to RNA (green) within the nucleus. The RNA travels to the cytoplasm where it is translated by ribosomes. Specific amino acids are associated with specific codons via tRNA (green). The final product of translation is complete, folded protein.

**SQ21. What if Beadle and Tatum analyzed the original irradiated haploid spores and did not analyze spores from the heterozygous strain. What information would they have missed?**

## III. RNA and an overview of gene expression (pp.238-240; pp.321-325)

We've established a connection between DNA and protein, but early on it was recognized that this connection must be indirect, at least in eukaryotes. Figure 2 illustrates why. If protein synthesis takes place in the cytoplasm while DNA is stuck in the nucleus, there must be something in between to mediate the transfer of genetic information. RNA is the obvious candidate for several reasons: (1) it is found in both the nucleus and the cytoplasm, (2) it is associated with ribosomes, (3) it looks like DNA.

Let's examine that last point.

**SQ22. How is RNA similar to DNA?**

**SQ23. Give three differences between RNA and DNA. (Think base composition, sugar composition, and usual structure).**

Figure 9-7 (p.240) gives a summary of the nucleotides found in RNA and DNA. They're the same, except for two things. First, RNA uses D-ribose instead of D-deoxyribose. This change has virtually zero effect on the properties of RNA as compared to DNA, except that proteins can use it to smell the difference between the two molecules. RNA is somewhat less stable than DNA, because the extra hydroxyl can attack the 3' carbon, leading to spontaneous breakdown. The second difference is that RNA uses uracil in place of thymine. The base-pairing properties of the two are identical. The big difference between the two in cellular organisms is that DNA is invariably double stranded, while RNA is invariably single stranded. The properties of DNA – double-stranded, stable – make it well suited as a depository of genetic information. The properties of RNA – single-stranded, less stable – make it well suited as a transient conveyer of information. We'll see why single-strandedness is useful next week, when we discuss the process of translation.

Figure 12-4 (p.322) illustrates what became known as the Central Dogma, that information goes from DNA to RNA to protein, but not the reverse. We know now that

the first step *can* be reversed. Some viruses (like HIV) can take RNA back to DNA. In cellular organisms, however, the generality holds for the most part.

The reason why the Overview of Gene Expression section exists in the text is Figure 12-5. It provides a concise overview of many of the processes we'll discuss in the coming weeks and connects them in a satisfying fashion. The process by which DNA is transcribed to RNA requires a binding site for RNA polymerase, the promoter. The decision of whether to transcribe the DNA or not rests with a regulatory sequence near the promoter. Both the promoter and regulatory sequence are no more than particular sequences of nucleotides that are recognized by protein. There is another particular nucleotide sequence later on (past the gene that is not shown), that signals RNA polymerase to stop.

In analogous fashion, the process by which RNA is translated to protein requires a binding site for the ribosome and a nearby sequence, the start codon, that determines where the ribosome will begin translation. At the end of the gene is a particular nucleotide sequence, the stop codon, that signals the ribosome to fall off the RNA.

**SQ24. The nucleotide sequence of DNA contains the information required to encode protein. What else does it contain?**