

## FROM GENES TO PROTEINS

So, what do genes actually do? How do they work? The original ideas have undergone some revisions. If you remember, Gregor Mendel published the results of his studies on pea plants in the *Proceedings of the Natural Science Society of Brno* in 1866. His paper was immediately forgotten until Hugo de Vries (Dutch Botanist), Karl Erich Correns (German Botanist), and Erich Tschermak von Seysenegg (Austrian Botanist) independent of each other, came to Mendel's same conclusions. Recognizing Mendel's contributions, the scientists cited Mendel's work in their papers. Thus Mendel is considered the father of genetics.

Not much later, Thomas Hunt Morgan, working with fruit flies, made the discovery that genes are found on the chromosomes and each gene must reside on a particular chromosome. This concept was not understood until Morgan's two papers, one in 1910 and the other in 1911, suggested the correlation.

Morgan's research, did not, however determine which components of the chromosome is the genetic code. Chromosomes are composed of two substances: DNA and histone proteins. For quite a while, no one knew which was the genetic code. Proteins had the inside track. Proteins are composed of amino acids and as you know, there are 20 common amino acids. DNA is composed of phosphate groups, ribose sugars, and four bases: adenine, thymine, cytosine and guanine. The idea at the time was the 4 bases of DNA could not code for enough information for the human genome. Instead, it looked as though the 20 amino acids had a better chance. It took several scientists many years to prove conclusively DNA, and not proteins, is the genetic code.

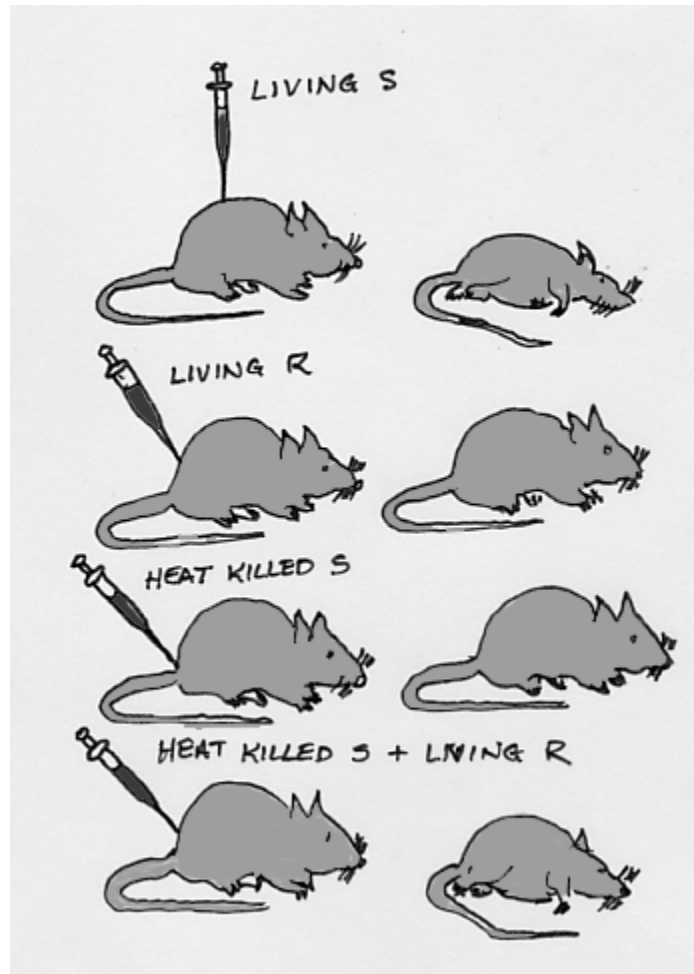
In 1909, the British physician Archibald Garrod said genes dictate phenotypes through enzymes that catalyze specific chemical processes in the cell. So far, we know genes are found on chromosomes and that phenotypes may be stipulated through enzymes.

In 1928, Frederick Griffith, a British medical officer, performed his famous experiment with mice. Griffith was studying the bacterium *Streptococcus pneumoniae*, the causative agent of bacterial pneumonia. Remember, there are several types of pneumonia: bacterial, viral, and protozoan (pneumocystis pneumonia). His work with *Streptococcus pneumoniae* involved two strains of the bacterium: rough and smooth. The smooth strain (S) was the pathogenic strain and caused pneumonia. The rough strain (R) was harmless. You could tell the difference by the way bacteria grew on Petri dishes. The smooth strain produced very neat, rounded colonies. The rough strain produced ragged edges to the colonies.

When Griffith injected healthy mice with the smooth (pathogenic) strain, the mice died – no surprise here. When he injected healthy mice with the rough strain, nothing happened. Again, no surprise since the rough strain is not pathogenic. These were his controls for his experiment.

## FROM GENES TO PROTEINS

In the next stage of his experiment, he heat killed the smooth strain and injected that into the mice. They lived. This, of course, was due to killing the pathogenic bacteria. However, it was the next process that really caused surprise in the research community. Griffith mixed heat killed S with live R. This was then injected into healthy mice and the mice died.



Griffith theorized that some transforming agent in the heat killed strain (S) had changed the nonpathogenic R strain into a pathogenic strain. The question was, what was the transforming agent?

Oswald Avery, Maclyn McCarty and Colin MacLeod, in 1944, stated the transforming agent was DNA. Saying it and proving it are two different things. It was not until 1952 that Martha Chase and Alfred Hershey definitively proved the transforming agent was indeed DNA.

## FROM GENES TO PROTEINS

### Chase and Hershey's Experiment

How do you prove Avery *et al* were correct in stating DNA is the transforming agent? The answer is found in viruses. Viruses are fairly simple in their construction. A virus is little more than a piece of genetic information (either DNA or RNA) surrounded by a protective coat of protein. It also may have sugars or other carbohydrates associated with the protein coat. For a virus to show any signs of life, in this case, reproduction, they must infect a cell and use the cell's "machinery" to make copies of the virus. Chase and Hershey utilized a particular virus which infects bacteria to prove DNA was the transforming agent.

### Bacteriophages

Bacteriophages, or simply phages, are viruses which infect specific forms of bacteria. Chase and Hershey used a bacteriophage called the T4 phage. The T4 phage is a DNA virus that infects the bacterium *Escherichia coli*. The structure of the T4 phage is rather unusual. It consists of a protein coat surrounding the DNA, called the phage "head." Attached to the head is more protein that forms a tail and then more proteins that form the tail fibers. The T4 phage looks very similar to some lunar landing craft.



When a phage finds an *E. coli* bacterium, it attaches to the surface of the bacterial wall by the tail fibers. The tail of the virus then penetrates the bacterial wall and the DNA in the head enters the bacterium through the tail. The DNA of the virus then becomes incorporated into the DNA of the bacterium. Remember, bacterial DNA is single stranded, single copy, with non-histone proteins. Once the viral DNA is incorporated into the bacterial DNA, one of two things happens.

## FROM GENES TO PROTEINS

### **Latency**

One possible outcome in the infection is for nothing to happen. Very simply, the viral DNA stays incorporated with the bacterial DNA and no “copies” of the virus are made. The period of inactivity by the virus can vary from a few hours to many months, and possibly years. This period is called latency.

### **Lysis**

Another possibility may occur immediately or when the virus ends its latency period. At this time, the virus uses the bacterial machinery to make copies of viral DNA and proteins. The DNA and proteins self assemble into the phage shape of head, tail and tail fibers. This occurs over and over again until the bacterium is literally bursting at the seams with phage particles. A critical mass is reached and the bacterium ruptures and viruses are released. These may, in turn, infect more bacteria.

### **Experimental Method of Chase and Hershey**

Remember, we cannot see bacteria or viruses with the unaided eye. As a consequence, Chase and Hershey had to devise some method by which events of transformation could be detected. They knew by this time DNA was composed of strands of nucleotides and histone proteins. DNA contains phosphates and proteins don't. Proteins, on the other hand, typically contain sulfur from specific “R” groups on the amino acids. DNA does not. Using this concept, Chase and Hershey grew viral particles with a source of radioactive sulfur. As the viruses reproduced, the copies incorporated radioactive sulfur into the protein coats of the viruses. These were allowed to infect bacteria. The reproductive cycle of the viruses was stopped at this point and the viruses and bacteria were placed in a blender. The blender ripped the protein coats from the bacteria but did not destroy the bacteria. The entire solution of bacteria and viruses were then centrifuged to separate the protein coats from the bacteria. *If* protein was the transforming agent, then the supernatant (the liquid) would have contained DNA and the bacteria would be radioactive with sulfur from proteins within the bacterial cells. Chase and Hershey found the supernatant contained the radioactive sulfur.

This did not completely prove DNA as the transforming agent. It simply proved proteins were not. What they did next did. They cultured a new batch of viruses with radioactive phosphorus. As the viruses reproduced, they incorporated the radioactive phosphorus *only* in the DNA. These were allowed to infect bacteria. Again, the bacteria and viruses were centrifuged to break the protein coats from the bacteria. The liquid was not found to be radioactive this time. Instead, the bacteria were radioactive with phosphorus. This conclusively proved DNA had been injected into the bacteria and not protein.

## FROM GENES TO PROTEINS

From George Beadle's and Boris Ephrussi's research in 1935, we understood that, in fruit flies, that mutations in eye color of the flies could be tied to no production of an enzyme that catalyzes the production of a pigment essential for normal eye color. In 1948, Beadle and Edward Tatum showed this relationship again in a fungus called *Neurospora crassa*, a form of bread mold.

This led to the idea that genes code for enzymes (proteins). The concept became one gene codes for one enzyme. Today, we realize that genes code, more specifically, for polypeptides (which may form proteins).

How do we get from genes to polypeptides and then to proteins?

### **Ribonucleic Acid (RNA)**

Hopefully you remember there are three major forms of RNA in the cell: messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA). All three have their origin in the nucleus of eukaryotic cells and the protoplasm of prokaryotic cells. The discussion that follows is typical of eukaryotic cells, but the process is similar, with variations, in prokaryotic cells.

### **Messenger RNA**

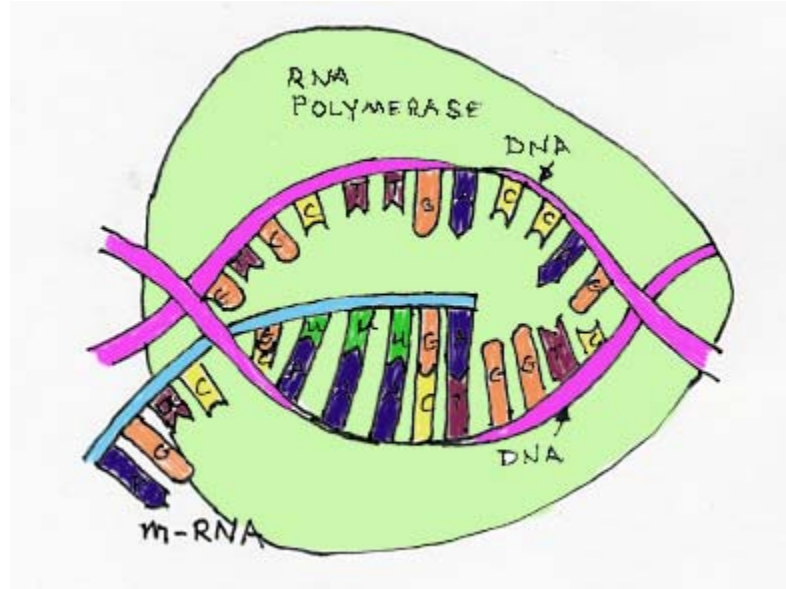
A segment of a single strand of DNA may constitute a gene. The double strand of DNA opens up and forms mRNA which "transcribes" the genetic message of the DNA by forming mRNA. As you may remember, RNA is similar to DNA except that (1) RNA is typically single stranded where DNA is typically double stranded (2) RNA contains a ribose sugar with one more oxygen than DNA's ribose sugar and (3) RNA does not contain the base thymine – instead it substitutes the base uracil.

So, how do we get the genetic message from DNA onto mRNA? Transcription to RNA is somewhat similar to replication of DNA by the semiconservative theory of replication. However, instead of opening up the entire strand of DNA for replication, transcription involves opening up only a small segment of the DNA molecule – that part which has the gene for what the cell needs.

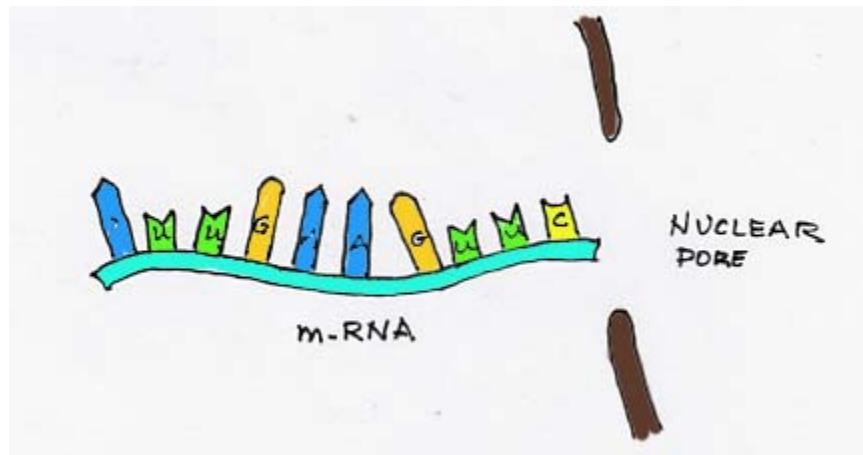
### **Transcription**

A series of enzymes opens up the segment of DNA, stabilizes it and prevents it from recoiling into the alpha helix until transcription is complete. Another enzyme, RNA polymerase, begins to form a mRNA nucleotide strand complementary to the DNA strand. This new mRNA does not join with the DNA. Instead, it is released into the nucleoplasm where it is modified for the next stage of the process.

## FROM GENES TO PROTEINS

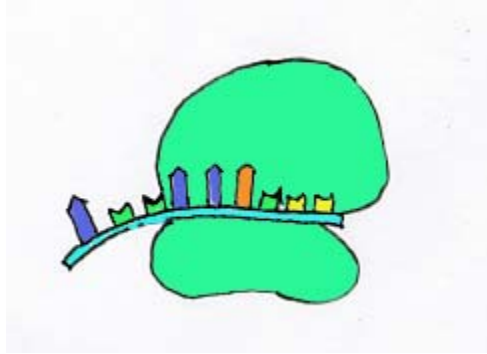


The modified mRNA now passes through the nuclear pores into the cytoplasm. This mRNA is a complimentary copy of the gene of the DNA. By complimentary, remember that if the DNA strand has a base that is cytosine, the complimentary base on the RNA would be guanine.



## FROM GENES TO PROTEINS

In the cytoplasm, ribosomal RNA (rRNA) finds the mRNA and attaches. Ribosomal RNA is primarily RNA but there are proteins also associated with it. Ribosomal RNA is composed of two subunits, one larger than the other. The two subunits migrate to the mRNA and attach to each other, “trapping” mRNA between them.



### **Translation**

How do we translate the genetic code into polypeptides (and thus proteins)? One first needs to know the genetic code. The research Marshall Nirenberg in 1966 and the confirmation of Nirenberg’s research by Gobind Khorana with further elaboration by Khorana, described the genetic code. In 1968, Robert Holley, Nirenberg, and Khorana were all three awarded the Nobel Prize in Medicine for their contributions to understanding the genetic code.

### **Genetic Code**

A sequence of three bases in a DNA molecule constitutes a codon. Codons code for amino acids. It was Nirenberg and Khorana that determined which amino acids. Here’s what they found. Remember, if the DNA codon reads ACA, the mRNA should read UGU.

## FROM GENES TO PROTEINS

Second Base

First base	Second Base				Third base
	U	C	A	G	
U	Phenylalanine	Serine	Tyrosine	Cysteine	U
	Phenylalanine	Serine	Tyrosine	Cysteine	C
	Leucine	Serine	STOP	STOP	A
	Leucine	Serine	STOP	Tryptophan	G
C	Leucine	Proline	Histidine	Arginine	U
	Leucine	Proline	Histidine	Arginine	C
	Leucine	Proline	Glutamine	Arginine	A
	Leucine	Proline	Glutamine	Arginine	G
A	Isoleucine	Threonine	Asparagine	Serine	U
	Isoleucine	Threonine	Asparagine	Serine	C
	Isoleucine	Threonine	Lysine	Arginine	A
	Methionine (START)	Threonine	Lysine	Arginine	G
G	Valine	Alanine	Aspartate	Glycine	U
	Valine	Alanine	Aspartate	Glycine	C
	Valine	Alanine	Glutamate	Glycine	A
	Valine	Alanine	Glutamate	Glycine	G

It should be immediately obvious that different codons may code for the same amino acid. For example, the amino acid valine may be coded for by GUU, GUC, GUA, and GUG. This redundancy is referred to as the degeneracy of the genetic code. The explanation is simple.

If a single base in DNA made a codon which codes for, the maximum number of amino acids that could be coded by those 4 bases would be  $4^1$  or 4. If two bases in DNA code for an amino acid, there would be  $4^2$  or 16 possible amino acids – not enough since there are 20 amino acids. However, if three bases in DNA code for an amino acid – which it does – there would be  $4^3$  or 64 possibilities of amino acids. That's 44 more than is actually needed. Mother Nature therefore builds in some redundancy to the genetic code.

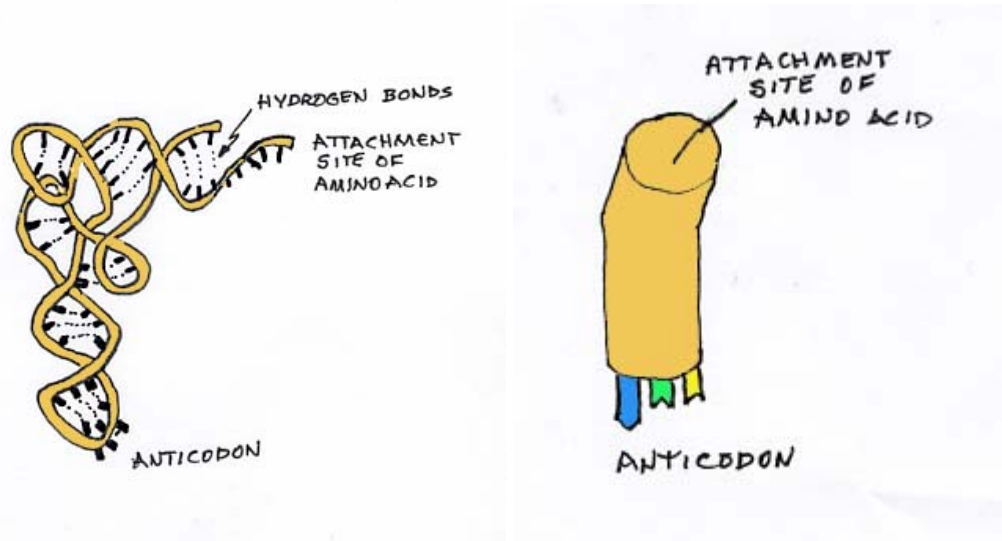
A second observation of the table will lead you to the start and stop codons. The genetic code is a good grammarian. It begins the gene with a “start here” codon, and then it tells the mRNA when to “stop” making the polypeptide.

Perhaps the most fascinating thing about the genetic code is its universality. By that, scientists mean that the codon UCU codes for Serine in dogs, cats, humans, plants, bacteria, fungi, protists, and even viruses. UCU codes for serine in all life forms on earth!

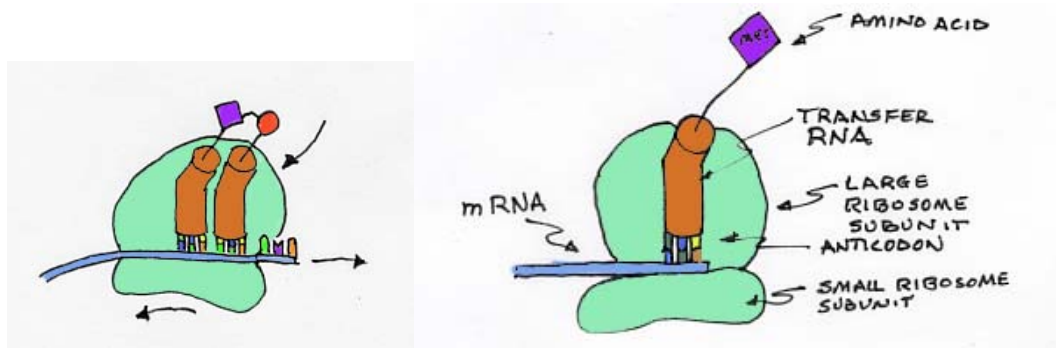
## FROM GENES TO PROTEINS

### Back to Translation

Previous to the discussion of the genetic code and the table, we left the mRNA attached to the rRNA by the two subunits of rRNA. How do we go from mRNA to amino acids to the polypeptide? The answer lies with tRNA. Transfer RNA is free in the cytoplasm of the cell. The purpose of tRNA is to search the cytoplasm for free amino acids. Each tRNA molecule has a complimentary codon (called the anticodon) to mRNA's codon. For example, if the mRNA sequence reads UCC, the tRNA anticodon would read AGG. That tRNA would attach only to the free roaming amino acid serine.



Once it finds serine in the cytoplasm, tRNA attaches to serine and carries it to the ribosome and mRNA where it matches the complimentary mRNA. The ribosome slips to the next codon on the mRNA, let's say AAA. Another tRNA with the anticodon to that (UUU) finds the amino acid lysine. It attaches to lysine and carries lysine to the ribosome and mRNA. We now have two amino acids attached to one ribosome.



## FROM GENES TO PROTEINS

The two amino acids are joined together by enzymes with the removal of water by a process called condensation synthesis (or dehydration synthesis) to form a dipeptide. The first tRNA is now freed to go back to the cytoplasm. The ribosome shift down one more codon, and the process repeats with tRNA, mRNA, and rRNA until we form a tripeptide, then a tetrapeptide, then a pentapeptide, etc. Eventually, the ribosome reaches the stop codon and the polypeptide is completed. You should refer to the chapter on cells to understand how the polypeptide is processed by the endoplasmic reticulum, sent to the Golgi apparatus, and then directed to its final destination in the cell.

### Expression of the Gene

In 1965, François Jacob and Jacques Monod, were awarded the Nobel Prize in Medicine for their work on how bacteria express the genetic code. Although their work was with bacteria, many of their concepts on gene expression hold for eukaryotic cells. The basis for gene expression is the operon.

### Operons

To understand operons, you must understand that if an organism does not need a protein, there is no reason for that organism to manufacture it. In the case of *Escherichia coli* (*E. coli*) found in your intestines, *E. coli* can use lactose (milk sugar) as a source of energy by breaking lactose down into glucose and galactose, two monosaccharides.

The reaction that breaks lactose into glucose and galactose (a hydrolysis reaction) is catalyzed by the enzyme  $\beta$ -galactosidase. Jacob and Monod found that if no lactose was present in the agar used to grow *E. coli*, then the bacterium would not produce the enzyme. Only when the bacterium was grown in the presence of lactose did the organism make  $\beta$ -galactosidase to break down lactose.

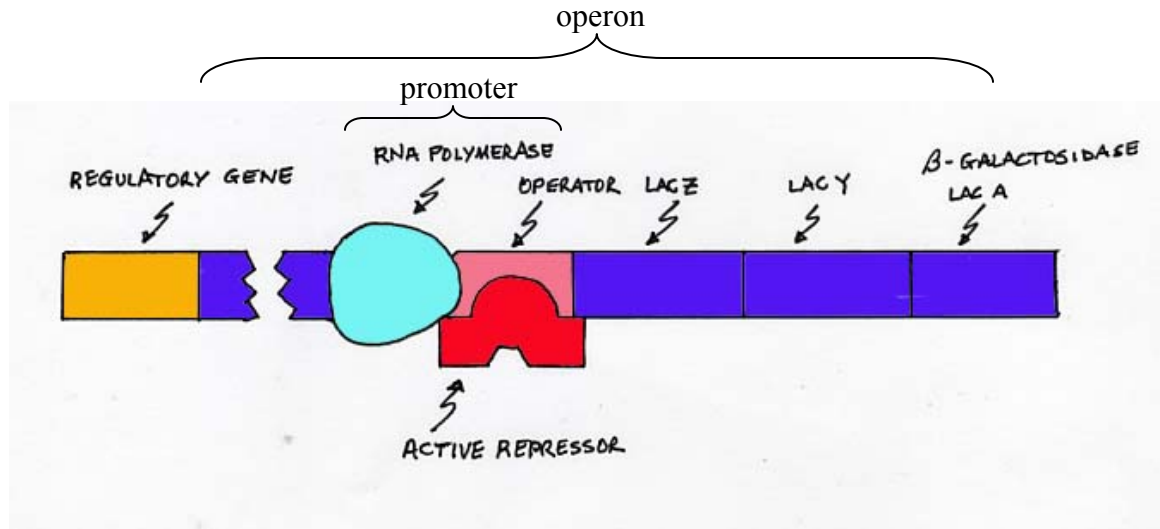
The control of manufacture of  $\beta$ -galactosidase is governed by several small contiguous sections of DNA called the operon. In the case of the production of the enzyme  $\beta$ -galactosidase, it is called the lac operon.

The gene for the production of  $\beta$ -galactosidase is located on a segment of a DNA molecule next to two other genes. It is not unusual that genes are grouped together and under the control of one operon. We are interested only in the lacA gene which makes  $\beta$ -galactosidase.

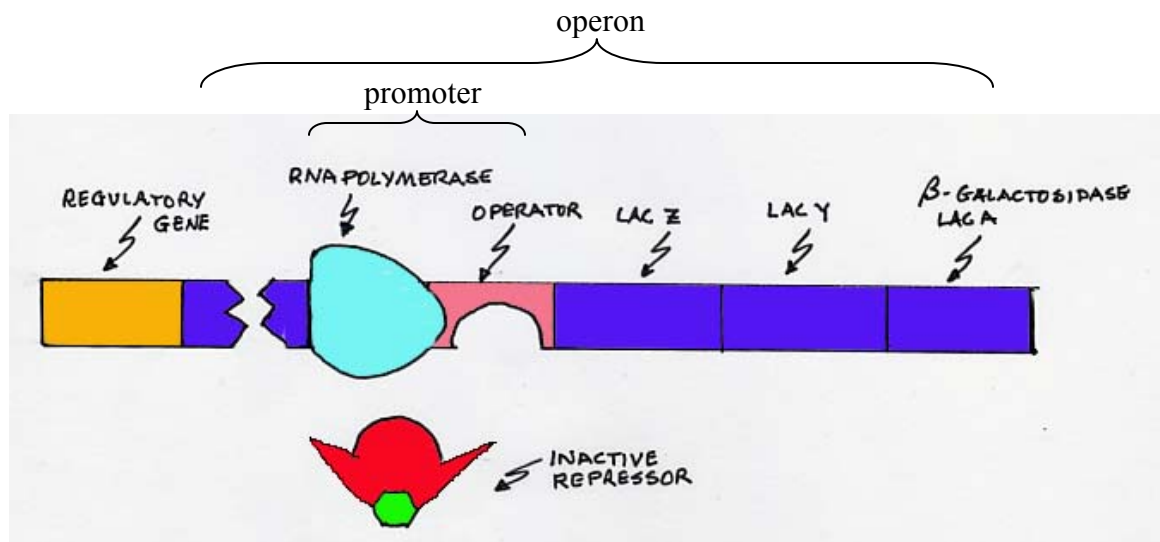
In front of these three genes is a region of DNA called the operator. In front of the operator is a segment of DNA that allows RNA polymerase (the enzyme necessary to form mRNA) to attach. RNA polymerase, when not physically blocked, slides down the segment of the DNA to the  $\beta$ -galactosidase gene (lacA) and produces mRNA to make the enzyme. The segments of DNA that include the genes, the operator and the portion of DNA that RNA polymerase attaches is called the operon.

## FROM GENES TO PROTEINS

Upstream from the operon is a regulatory gene. When lactose is not present in the agar, the bacterium has no need to produce  $\beta$ -galactosidase. As a result, the regulatory gene produces a protein called the active repressor. This repressor attaches to the operator. When the active repressor attaches to the operator, RNA polymerase cannot move any further and its path to the genes are blocked. No  $\beta$ -galactosidase is produced.



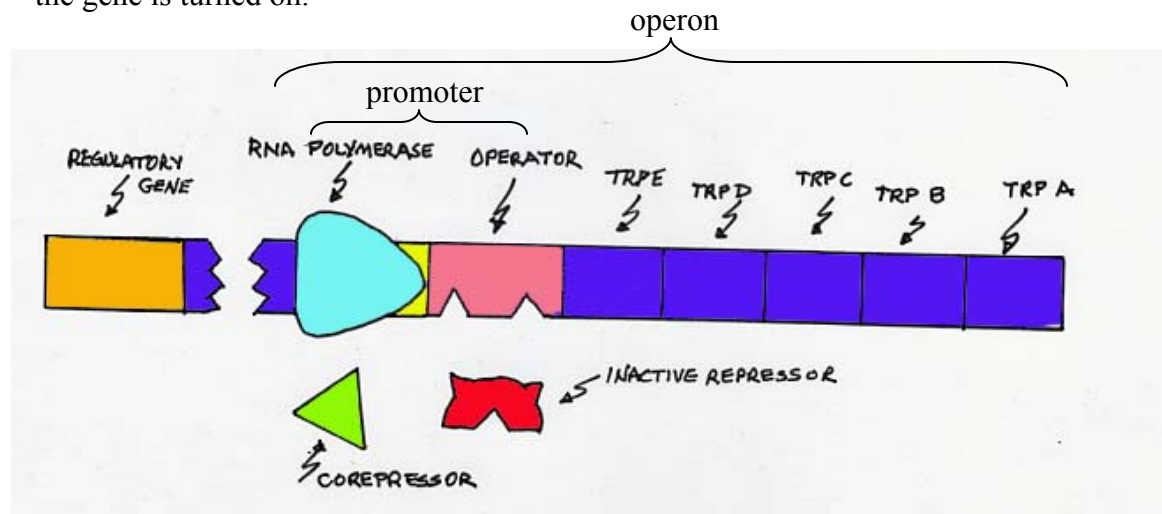
However, if lactose is present in the agar, then  $\beta$ -galactosidase is needed. The bacterium absorbs lactose which in turn reacts with the repressor to deactivate it and prevent the repressor from interacting with the operator. Since the pathway to the gene is free and not blocked, RNA polymerase slips down the DNA and produces mRNA and  $\beta$ -galactosidase is manufactured. Lactose, in this case, serves as an *inducer* to inactivate the repressor. The lac operon is called an inducible operon.



## FROM GENES TO PROTEINS

Jacob and Monod found another mechanism to turn genes on and off in bacteria. This method is called a repressible operon. If *E. coli* is grown in agar that contains the amino acid tryptophan, the gene for manufacturing the enzyme necessary for the production of tryptophan is turned off. If you grow *E. coli* without tryptophan, then the bacterium needs the enzyme for the production of the amino acid. The gene is thus turned on.

In this repressible operon, tryptophan itself serves, not to deactivate the repressor as in the lac operon, but instead to activate the repressor. When the active repressor attaches to the operator, the gene is turned off and tryptophan is not manufactured. If tryptophan is not present in the agar, then the repressor becomes deactivated and the operator site is free for RNA polymerase to make the enzyme – the gene is turned on.



It's through Jacob and Monod's models of turning genes on and off that scientists began to understand how human cells turn genes on and off. This, in turn, has led to discoveries in how to treat cancerous cells and genetic defects.

### Genetic Defects

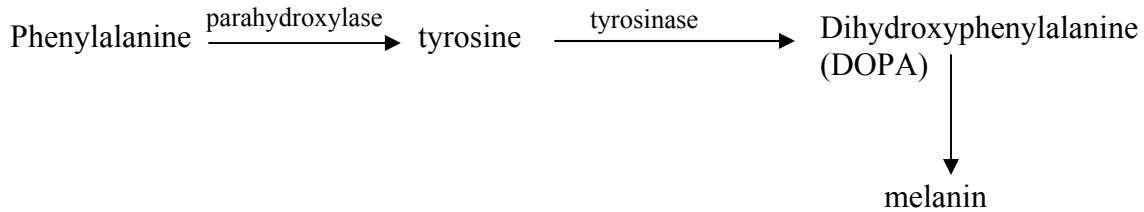
What happens when genes go awry? Depending on the importance of the gene, the results could be severe, even devastating. In other cases, there may be less severe consequences.

### Albinism

Albinism in animals (and humans) occurs when the production of the pigment melanin is blocked. Melanin is one of the three major aspects of skin, hair and eye color. (The other two are the amount of carotene present and the amount blood vessels which show through.)

## FROM GENES TO PROTEINS

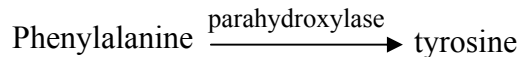
Melanin production occurs through a series of biochemical reactions in the body, each catalyzed by enzymes. Melanin production involves the amino acids phenylalanine and tyrosine. Phenylalanine is one of the essential amino acids which must be obtained through your diet. Albinism is related to tyrosine production. To produce melanin, phenylalanine must be converted to tyrosine in the presence of the enzyme parahydroxylase. Tyrosine must be converted to dihydroxyphenylalanine (DOPA). DOPA is converted to melanin. The enzyme tyrosinase catalyzes the conversion of tyrosine to DOPA. Block tyrosinase and no DOPA is produced, and thus, no melanin – the result is albinism.



### Phenylketonuria (PKU)

Another example of phenylalanine's importance in the body is PKU. PKU occurs in approximately 1 out of every 10,000 to 15,000 births each year in the U.S. Those children with the disorder cannot break down phenylalanine. Phenylalanine and its byproduct, phenylpyruvate, accumulate in the blood to toxic levels. Phenylalanine concentrations can reach the point that the child's urine contains high levels of the amino acid. Tragically, toxic levels result in irreversible mental retardation of the infant. The good news is that PKU can be detected at birth and the child can be placed on a diet low in phenylalanine (Read the label on many diet soft drinks – it says not for phenylketonurics.). Screening, however, must be done at birth.

PKU occurs when there is a defect in the gene for the production of the enzyme parahydroxylase. Remember, in melanin production, this enzyme is essential in converted phenylalanine into tyrosine. If the enzyme is not produced, phenylalanine is not converted to tyrosine and phenylalanine and phenylpyruvic acid levels build to toxic levels with the result of PKU.



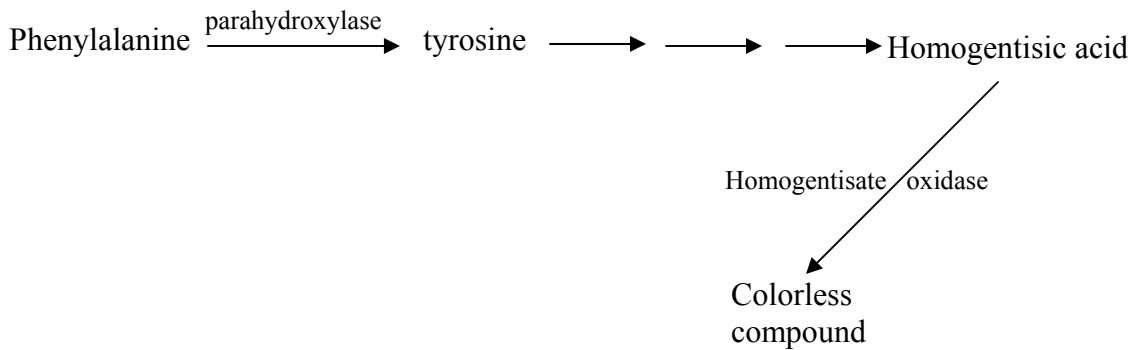
## FROM GENES TO PROTEINS

### Alkaptonuria

Alkaptonuria is a condition where the cartilage in the body becomes darkened. This is particularly noticeable in the ears, wrist, and elbows. When people reach middle or old age, the sites of darkening have been related to arthritis. One of the more unusual effects of alkaptonuria is that individuals with the disorder produce a urine which turns black upon exposure to air. This phenomenon is tied to the production of homogentisic acid.

Normally, people have an enzyme, homogentisate oxidase, which breaks homogentisic acid down into a colorless compound - no homogentisate oxidase, no breakdown of homogentisic acid – thus the black urine.

Again, the reaction begins with phenylalanine. Phenylalanine is broken down into tyrosine (by parahydroxylase). Tyrosine, through a series of reactions, may be converted to homogentisic acid. This in turn, should be broken down into a colorless compound by the enzyme homogentisate oxidase.



### Enzyme Deficiencies

The above three disorders are tied to a defect in the genes to produce specific enzymes. To a degree, the results are either mild (alkaptonuria and albinism) to potentially devastating (PKU). The question is, how are the genes defective? The genes have mutated.

### Mutations

Mutations are changes in genes of cells. We previously looked at one form of mutations called nondisjunction of chromosomes. There is another type of mutation that affects a single base in the DNA of the gene. This type of mutation is called a point mutation.

## FROM GENES TO PROTEINS

### Point Mutations

There are two basic types of point mutations (1) substitutions and (2) insertions and deletions.

### Substitutions

This occurs when a single nucleotide (and its complement) are replaced by another nucleotide and its complement. For example, say the nucleotide contains the base adenine. Thymine normally pairs opposite adenine in a DNA molecule. Instead of adenine, substitute the base guanine. Now the thymine doesn't work. Instead, in replicating the DNA, cytosine must be inserted at this point.

Some substitutions are called silent mutations. Since there is redundancy in the genetic code (64 possible codons) if the base substitution takes place, it may not change the amino acid. For example, UUU codes for phenylalanine. If the last base is changed to a cytosine, the codon now reads UUC. This also codes for phenylalanine.

Many of substitution mutations are missense mutations. The mutation codes for an amino acid, just not the right amino acid. Normally, the change in a single amino acid in a polypeptide is not significant. However, suppose the mutation results in a stop codon. The polypeptide may be cut short and that has a dramatic effect on the production of the protein. This "stop" mutation is called a nonsense mutation.

### Insertions and Deletions

These types of mutations occur when there are additions or deletions of nucleotide pairs in genes. These often have disastrous effects on protein production, and thus to the individual to whom they occur. Remember, mRNA is produced by RNA polymerase by reading a series of three bases (codon) at a time. Most of these codons must be read in order.

If a segment of a DNA molecule reads UUU CCG UCC AAA and we insert a base G after the third U, the DNA molecule now reads UUU GCC GUC CAA A. Since we are shifting the entire order of bases down, this is called a frameshift mutation. The same thing can occur with a deletion.

## FROM GENES TO PROTEINS

### **Types of Mutations**

Some mutations are simply spontaneous. They may occur during replication of DNA, the repair of DNA or the recombination of DNA in bacteria. These often lead to point mutations.

Other mutations may be caused by physical means. X-rays and UV radiation can cause mutations.

Chemical mutations occur (1) when a chemical affects the chemical structure of one of the bases (2) chemicals affect the DNA replication process by inserting themselves into the molecule or (3) chemicals mimic bases and result in an incorrect base pairing in DNA.

One way to determine the effect of chemicals to cause mutations is to run the Ames test on bacteria to see if mutations occur in bacteria. If this occurs, the chemical agent is suggested to be carcinogenic.