**How DNA Works by Craig Freudenrich, Ph.D. Source:** [**http://science.howstuffworks.com/life/cellular-microscopic/dna1.htm/printable**](http://science.howstuffworks.com/life/cellular-microscopic/dna1.htm/printable)

**Introduction to How DNA Works**

Like the one ring of power in Tolkien's "Lord of the Rings," **deoxyribonucleic acid** (DNA) is the master molecule of every [cell](http://science.howstuffworks.com/life/cellular-microscopic/cell.htm). It contains vital information that gets passed on to each successive generation. It coordinates the making of itself as well as other molecules (proteins). If it is changed slightly, serious consequences may result. If it is destroyed beyond repair, the cell dies.

Changes in the DNA of cells in multicellular organisms produce variations in the characteristics of a species. Over long periods of time, natural selection acts on these variations to [evolve](http://science.howstuffworks.com/life/evolution/evolution.htm) or change the species.

The presence or absence of DNA evidence at a [crime scene](http://science.howstuffworks.com/csi.htm) could mean the difference between a guilty verdict and an acquittal. DNA is so important that the United States government has spent enormous amounts of money to unravel the sequence of DNA in the human genome in hopes of understanding and finding cures for many genetic diseases. Finally, from the DNA of one cell, we can [clone](http://science.howstuffworks.com/life/genetic/cloning.htm) an animal, a plant or perhaps even a human being.

But what is DNA? Where is it found? What makes it so special? How does it work? In this article, we will look deep into the structure of DNA and explain how it makes itself and how it determines all of your traits. First, let's look at how DNA was discovered.

DNA is one of a class of molecules called **nucleic acids**. Nucleic acids were originally discovered in 1868 by Friedrich Meischer, a Swiss biologist, who isolated DNA from pus cells on bandages. Although Meischer suspected that nucleic acids might contain genetic information, he could not confirm it.

In 1943, Oswald Avery and colleagues at Rockefeller University showed that DNA taken from a bacterium, *Streptococcus pneumonia*, could make non-infectious bacteria become infectious. These results indicated that DNA was the information-containing molecule in the cell. The information role of DNA was further supported in 1952 when Alfred Hershey and Martha Chase demonstrated that to make new viruses, a **bacteriophage** virus injected DNA, not protein, into the host cell.

So scientists had theorized about the informational role of DNA for a long time, but nobody knew how this information was encoded and transmitted. Many scientists guessed that the structure of the molecule was important to this process. In 1953, James D. Watson and Francis Crick discovered the structure of DNA at Cambridge University. The story was described in James Watson's book "The Double Helix" and brought to the screen in the movie, "The Race for the Double Helix." Basically, Watson and Crick used molecular modeling techniques and data from other investigators (including Maurice Wilkins, Rosalind Franklin, Erwin Chargaff and Linus Pauling) to solve the structure of DNA. Watson, Crick and Wilkins received the Nobel Prize in Medicine for the discovery of DNA's structure (Franklin, who was Wilkins' collaborator and provided a key piece of data that revealed the structure to Watson and Crick, died before the prize was awarded).

**DNA Structure**

DNA is o­ne of the **nucleic acids**, information-containing molecules in the cell (**ribonucleic acid**, or RNA, is the other nucleic acid). DNA is found in the nucleus of every human cell. (See the sidebar at the bottom of the page for more about RNA and different types of cells). The information in DNA:

* guides the cell (along with RNA) in making new proteins that determine all of our biological traits
* gets passed (copied) from one generation to the next

The key to all of these functions is found in the molecular structure of DNA, as described by Watson and Crick.

Although it may look complicated, the DNA in a cell is really just a pattern made up of four different parts called **nucleotides**. Imagine a set of blocks that has only four shapes, or an alphabet that has only four letters. DNA is a long string of these blocks or letters. Each nucleotide consists of a sugar (**deoxyribose**) bound on one side to a **phosphate group** and bound on the other side to a **nitrogenous base**.

There are two classes of nitrogen bases called **purines** (double-ringed structures) and **pyrimidines** (single-ringed structures). The four bases in DNA's alphabet are:

* **adenine (A)** - a purine
* **cytosine(C)** - a pyrimidine
* **guanine (G)** - a purine
* **thymine (T)** - a pyrimidine

**Strands of DNA are made of the sugar and phosphate portions of the nucleotides, while the middle parts are made of the nitrogenous bases. The nitrogenous bases on the two strands of DNA pair up, purine with pyrimidine (A with T, G with C), and are held together by weak hydrogen bonds.**

Watson and Crick discovered that DNA had two sides, or strands, and that these strands were twisted together like a twisted ladder -- the **double helix**. The sides of the ladder comprise the sugar-phosphate portions of adjacent nucleotides bonded together. The phosphate of one nucleotide is **covalently bound** (a bond in which one or more pairs of electrons are shared by two atoms) to the sugar of the next nucleotide. The hydrogen bonds between phosphates cause the DNA strand to twist. The nitrogenous bases point inward on the ladder and form pairs with bases on the other side, like rungs. Each base pair is formed from two complementary nucleotides (purine with pyrimidine) bound together by hydrogen bonds. The base pairs in DNA are **adenine with thymine** and **cytosine with guanine**.

**DNA has a spiral staircase-like structure. The steps are formed by the nitrogen bases of the nucleotides where adenine pairs with thymine and cytosine with guanine.**

In the next section we'll find out how long DNA strands fit inside a tiny cell.

**HYDROGEN BOND**

A hydrogen bond is a **weak chemical bond** that occurs between hydrogen atoms and more electronegative atoms, like oxygen, nitrogen and fluorine. The participating atoms can be located on the same molecule (adjacent nucleotides) or on different molecules (adjacent nucleotides on different DNA strands). Hydrogen bonds do not involve the exchange or sharing of electrons like covalent and ionic bonds. The weak attraction is like that between the opposite poles of a magnet. Hydrogen bonds occur over short distances and can be easily formed and broken. They can also stabilize a molecule.

**Fitting Inside a Cell**

DNA is a long molecule. For example, a typical bacterium, like *E. coli*, has one DNA molecule with about 3,000 genes (A gene is a specific sequence of DNA nucleotides that codes for a protein. We'll talk about this later). If drawn out, this DNA molecule would be about 1 millimeter long. However, a typical *E. coli* is only 3 microns long (3 one-thousandths of a millimeter).So to fit inside the cell, the DNA is highly coiled and twisted into one circular chromosome.

Complex organisms, like plants and animals, have 50,000 to 100,000 genes on many different chromosomes (humans have 46 chromosomes). In the cells of these organisms, the DNA is twisted around bead-like proteins called **histones**. The histones are also coiled tightly to form chromosomes, which are located in the nucleus of the cell. When a cell reproduces, the chromosomes (DNA) get copied and distributed to each offspring, or daughter, cell. Non-sex cells have two copies of each chromosome that get copied and each daughter cell receives two copies (**mitosis**). During meiosis, precursor cells have two copies of each chromosome that gets copied and distributed equally to four sex cells. The sex cells (sperm and egg) have only one copy of each chromosome. When sperm and egg unite in fertilization, the offspring have two copies of each chromosome (see [How Sex Works](http://health.howstuffworks.com/pregnancy-and-parenting/pregnancy/conception/human-reproduction.htm)).

In the next section we'll look at how the DNA replication process works.

**The double helix of DNA unwinds and each side serves as a pattern to make a new molecule.**

Image courtesy [U.S. Department of Energy Human Genome Program](http://www.ornl.gov/)

**DNA Replication**

DNA carries the information for making all of the cell's proteins. These pro­teins implement all of the functions of a living organism and determine the organism'­s characteristics. When the cell reproduces, it has to pass all of this information on to the daughter cells.

Before a cell can reproduce, it must first **replicate**, or make a copy of, its DNA. Where DNA replication occurs depends upon whether the cells is a prokaryote or a eukaryote (see the RNA sidebar on the previous page for more about the types of cells). DNA replication occurs in the cytoplasm of prokaryotes and in the nucleus of eukaryotes. Regardless of where DNA replication occurs, the basic process is the same.

The structure of DNA lends itself easily to DNA replication. Each side of the double helix runs in opposite (**anti-parallel**) directions. The beauty of this structure is that it can unzip down the middle and each side can serve as a pattern or template for the other side (called **semi-conservative replication**). However, DNA does not unzip entirely. It unzips in a small area called a **replication fork**, which then moves down the entire length of the molecule.

Let's look at the details:

1. An enzyme called **DNA gyrase** makes a nick in the double helix and each side separates
2. An enzyme called **helicase** unwinds the double-stranded DNA
3. Several small proteins called **single strand binding proteins** (SSB) temporarily bind to each side and keep them separated
4. An enzyme complex called **DNA polymerase** "walks" down the DNA strands and adds new nucleotides to each strand. The nucleotides pair with the complementary nucleotides on the existing stand (A with T, G with C).
5. A subunit of the DNA polymerase **proofreads** the new DNA
6. An enzyme called **DNA ligase** seals up the fragments into one long continuous strand
7. The new copies **automatically wind up again**

Different types of cells replicated their DNA at different rates. Some cells constantly divide, like those in your hair and fingernails and bone marrow cells. Other cells go through several rounds of cell division and stop (including specialized cells, like those in your [brain](http://science.howstuffworks.com/life/inside-the-mind/human-brain/brain.htm), [muscle](http://health.howstuffworks.com/human-body/systems/musculoskeletal/muscle.htm) and [heart](http://health.howstuffworks.com/human-body/systems/circulatory/heart.htm)). Finally, some cells stop dividing, but can be induced to divide to repair injury (such as skin cells and liver cells). In cells that do not constantly divide, the cues for DNA replication/cell division come in the form of chemicals. These chemicals can come from other parts of the body (hormones) or from the environment.

**ANIMAL VS. PLANT DNA**

The DNA of all living organisms has the same structure and code, although some viruses use RNA as the information carrier instead of DNA. Most animals have two copies of each chromosome. In contrast, plants may have more than two copies of several chromosomes, which usually arise from errors in the distribution of the chromosomes during cell reproduction. In animals, this type of error usually causes genetic diseases that are usually fatal. For some unknown reasons, this type of error is not as devastating to plants.

**DNA contains the information to make proteins, which carry out all the functions and characteristics of living organisms.**

Photo courtesy [U.S. Department of Energy Genomics: GTL Program](http://genomicsgtl.energy.gov/)

**What DNA Does**

DNA carr­ies all of the information for your physical characteristics, which are essentially determined by proteins. So, DNA contains the instructions for making a protein. In DNA, each protein is encoded by a **gene** (a specific sequence of DNA nucleotides that specify how a single protein is to be made). Specifically, the order of nucleotides within a gene specifies the order and types of amino acids that must be put together to make a protein.

A protein is made of a long chain of chemicals called **amino acids** Proteins have many functions:

* **Enzymes** that carry out chemical reactions (such as digestive enzymes)
* **Structural proteins** that are building materials (such as collagen and nail keratin)
* **Transport proteins** that carry substances (such as oxygen-carrying hemoglobin in [blood](http://health.howstuffworks.com/human-body/systems/circulatory/blood.htm))
* **Contraction proteins** that cause muscles to compress (such as actin and myosin)
* **Storage proteins** that hold on to substances (such as albumin in egg whites and iron-storing ferritin in your spleen)
* **Hormones** - chemical messengers between cells (including [insulin](http://health.howstuffworks.com/medicine/medication), [estrogen](http://health.howstuffworks.com/wellness/women/menopause/how-menopause-works2.htm), [testosterone](http://people.howstuffworks.com/men.htm), cortisol, et cetera)
* **Protective proteins** - antibodies of the [immune system](http://health.howstuffworks.com/human-body/systems/immune/immune-system.htm), clotting proteins in blood
* **Toxins** - poisonous substances, (such as bee venom and [snake](http://animals.howstuffworks.com/snakes/snake.htm) venom)

The particular sequence of amino acids in the chain is what makes one protein different from another. This sequence is encoded in the DNA where one gene encodes for one protein.

**The genetic code consists of 3-base "words" or codons that specify particular amino acids. The order of the codons designates the order of the amino acids in the protein.**

Photo courtesy [U.S. Department of Energy Human Genome Program](http://www.ornl.gov/)

How does DNA encode the information for a protein? There are only four DNA bases, but there are 20 amino acids that can be used for proteins. So, groups of three nucleotides form a word (**codon**) that specifies which of the 20 amino acids goes into the protein (a 3-base codon yields 64 possible patterns (4\*4\*4), which is more than enough to specify 20 amino acids. Because there are 64 possible codons and only 20 amino acids, there is some repetition in the genetic code. Also, the order of codons in the gene specifies the order of amino acids in the protein. It may require anywhere from 100 to 1,000 codons (300 to 2,000 nucleotides) to specify a given protein. Each gene also has codons to designate the beginning (**start codon**) and end (**stop codon**) of the gene.

In the next few sections, we'll see how proteins are built.

**RNA (RIBONUCLEIC ACID)**

RNA is the **other** nucleic acid. It differs from DNA in three major ways:

* The sugar is **ribose** instead of deoxyribose
* There is only one strand instead of two
* RNA has **uracil (U)** instead of thymine. So, the base pairs in RNA are **cytosine with guanine** and **adenine with uracil**.

In a **prokaryotic cell** (one with no internal membrane-bound organelles like a bacterium), both DNA and RNA are found in the cytoplasm. In a **eukaryotic cell** (one with internal membrane-bound organelles, like humans), RNA can be found in the nucleus and cytoplasm, while DNA is only found in the nucleus.

**Building a Protein: Transcription**

Building proteins is very much like building a house:

* The master blueprint is DNA, which contains all of the information to build the new protein (house).
* The working copy of the master blueprint is called **messenger RNA** (mRNA), whic­h is copied from DNA.
* The construction site is either the **cytoplasm** in a prokaryote or the **endoplasmic reticulum** (ER) in a eukaryote.
* The building materials are amino acids.
* The construction workers are ribosomes and transfer RNA molecules.

Let's look at each phase of the new construction more closely.

In a eukaryote, DNA never leaves the nucleus, so its information must be copied. This copying process is called **transcription** and the copy is mRNA. Transcription takes place in the cytoplasm (prokaryote) or in the nucleus (eukaryote). The transcription is performed by an enzyme called RNA polymerase. To make mRNA, RNA polymerase:

1. Binds to the DNA strand at a specific sequence of the gene called a **promoter**
2. Unwinds and unlinks the two strands of DNA
3. Uses one of the DNA strands as a guide or template
4. Matches new nucleotides with their complements on the DNA strand (G with C, A with U -- remember that RNA has uracil (U) instead of thymine (T))
5. Binds these new RNA nucleotides together to form a complementary copy of the DNA strand (mRNA)
6. Stops when it encounters a termination sequence of bases (stop codon)

mRNA is happy to live in a single-stranded state (as opposed to DNA's desire to form complementary double-stranded helixes). In prokaryotes, all of the nucleotides in the mRNA are part of codons for the new protein. However, in eukaryotes only, there are extra sequences in the DNA and mRNA that don't code for proteins called **introns**. This mRNA is then further processed:

* Introns get cut out
* The coding sequences get spliced together
* A special nucleotide "cap" gets added to one end
* A long tail consisting of 100 to 200 adenine nucleotides is added to the other end

No one knows why this processing occurs in eukaryotes. Finally, at any one moment, many genes are being transcribed simultaneously according to the cell's needs for specific proteins.

The working copy of the blueprint (mRNA) must now go the construction site where the workers will build the new protein. If the cell is a prokaryote such as an *E. coli* bacterium, then the site is the cytoplasm. If the cell is a eukaryote, such as a human cell, then the mRNA leaves the nucleus through large holes in the nuclear membrane (nuclear pores) and goes to the endoplasmic reticulum (ER).

Next, we'll learn about translation -- the assembly process.

**This is a table of the genetic code based on mRNA codons. Some tables are based on the DNA codons.**

**Building a Protein: Translation**

To continue with our house example, once the working copy of the blueprint has reached the site, the workers must assemble the materials according to the instructio­ns; this process is called **translation**. In the case of a protein, the workers are the ribosomes and special RNA molecules called **transfer RNA (tRNA)**. The construction materials are the amino acids.

First, let's look at the **ribosome**. The ribosome is made of RNA called **ribosomal RNA** (rR­NA). In prokaryotes, rRNA is made in the cytoplasm; in eukaryotes, rRNA is made in the nucleolus. The ribosome has two parts, which bind on either side of the mRNA. Within the large part are two "rooms" (P and A sites) that will fit two adjacent codons of the mRNA, two tRNA molecules and two amino acids. At first, the P site holds the first codon in the mRNA and A site holds the next codon.

Next, let's look at the tRNA molecules. Each tRNA has a binding site for an amino acid. Because each tRNA is specific for a single amino acid, it must be able to recognize the codon on the mRNA that codes for that particular amino acid. Therefore, each tRNA has a specific three-nucleotide sequence called an anti-codon that matches up with the appropriate mRNA codon, like a lock and key. For example, if a codon on mRNA has the sequence **...-uracil-uracil-uracil-... (UUU)** which codes for the amino acid phenylalanine, then the anti-codon on the phenylalanine tRNA will be **adenine-adenine-adenine (AAA)**; remember that A binds with U in RNA. The tRNA molecules float in the cytoplasm and bind free amino acids. Once bound to amino acids, the tRNAs (also called amino-acyl tRNAs) will seek out ribosomes.

Finally, let's look at the events in the synthesis of new proteins. For example, let's consider a small mRNA molecule with the following sequence:

All mRNA molecules begin with AUG (the start codon). UGA, UAA, and UAG are stop codons; sto­p codons have no corresponding tRNA molecules (Actual mRNA molecules have hundreds of codons).

The corresponding sequence of tRNA anti-codons will be:

There is no tRNA corresponding to the stop codons.

The amino acid sequence specified by this small mRNA is:

We know this sequence of amino acids by using a table of the genetic code. The genetic code table below is for mRNA and specifies the bases in the first, second and third positions of the codon with their corresponding amino acids.

Let's read the amino acid specified by the mRNA codon, AUG. First, place your left finger on the first position codon (A), in the first column of the table. Move your left finger across the row under the second position codon (U) in the first row. Now, place your right finger over the third position codon (G) in the same row of the last column (G). Move your right finger across the row until it meets your left finger and read the amino acid (methionine).

**MITOCHONDRIAL DNA**

Mitochondria (and chloroplasts in plants) have their own small loops of DNA and reproduce on their own, independent of what goes on in the nucleus. This DNA codes for some mitochondrial proteins, but others are provided from the DNA stored in the nucleus. Mitochondria resemble an early form of bacteria, which is thought to have been captured into eukaryotic cells early in the history of life on Earth. The bacteria coexisted with the cell (endosymbiosis) and evolved into mitochondria. Another unique aspect of mitochondrial DNA is that you inherit it only from your mother (the mitochondria that exist in the egg cell). Although the sperm cell that fertilizes the egg contains a mitochondrion from the father, it does not get released and passed on.

In the next section we'll look at the protein synthesis process.

**The Protein Synthesis Process**

Now let's look at the order of events in the synthesis of our protein from our sample mRNA:

1. A ribosome binds to mRNA with the AUG codon in the P-site and the UUU codon in the A-site.
2. An amino acyl-tRNA (anti-codon = UAC) with an attached methionine comes into the P-site of the ribosome
3. An amino acyl-tRNA (anti-codon = AAA) with an attached phenylalanine comes into the A-site of the ribosome
4. A chemical bond forms between the methionine and phenylalanine (in a protein, this covalent bond is called a**peptide bond**).
5. The methionine-specific tRNA leaves the P-site and goes off to gather another methionine
6. The ribosome shifts so that the P-site now contains the UUU codon with the attached phenyl-alanine tRNA and the next codon (ACA) now occupies the A-site.
7. An amino acyl-tRNA (anti-codon) with an attached threonine comes into the A-site of the ribosome.
8. A peptide bond forms between the phenylalanine and the threonine.
9. The phenylalanine-specific tRNA leaves the P-site and goes off to find another phenylalanine.
10. The ribosome shifts down one codon so that the stop sequence is now in the A-site. Upon encountering the stop sequence, the ribosome detaches from the mRNA and splits into its two parts, the threonine-specific tRNA releases its threonine and leaves, and the new protein floats away

Several ribosomes can attach to a molecule of mRNA one after another and begin making proteins. So several proteins can be made from one mRNA. In fact, in *E. coli* bacteria, translation of the mRNA begins even before transcription is finished.

In the next section we'll look at DNA mutation.

**Variations in the sequence of genes during can have important consequences and cause disease.**

Photo courtesy [U.S. Department of Energy Human Genome Program](http://www.ornl.gov/)

**DNA Mutation, Variation and Sequencing**

In the h­uman genome, there are 50,000 to 100,000 genes. As DNA polymerase copies the DNA sequence, some mistakes occur. For example, one DNA base in a gene might get substituted for another. This is called a **mutation** (specifically a **point mutation**) or variation in the gene. Because the genetic code has built-in redundancies, this mistake might not have much effect on the protein made by the [gene](http://science.howstuffworks.com/life/genetic/gene-pool.htm). In some cases, the error might be in the third base of a codon and still specify the same amino acid in the protein. In other cases, it may be elsewhere in the codon and specify a different amino acid. If the changed amino acid is not in a crucial part of the protein, then there may be no adverse effect. However, if the changed amino acid is in a crucial part of the protein, then the protein may be defective and not work as well or at all; this type of change can lead to disease.

Other types of mutations in DNA can occur when small segments of DNA break off the chromosome. These segments can get placed back at another spot in the chromosome and interrupt the normal flow of information. These types of mutations (deletions, insertions, inversions) usually have severe consequences.

As noted above, there is lots of extra DNA in the human genome that does not code for proteins. What this extra non-coding DNA does is actively being researched. Perhaps some of it is merely spacing to hold the genes a certain distance apart for the enzymes of transcription. Some might be places where environmental chemicals might bind and affect DNA transcription and/or translation. Also, within this extra DNA, there are many variation sequences that are used in DNA typing (see [How DNA Evidence Works](http://science.howstuffworks.com/life/genetic/dna-evidence.htm)).

**DNA Sequencing**

The Human Genome Project (HGP) was initiated in the 1990s with the goal of determining the sequence of the entire human genome. What genes were present? Where they were located? What were the sequences of the genes and the intervening DNA (non-coding DNA)? This task was monumental, along the order of the US Apollo Project to place a man on the Moon. The HGP scientists and contractors developed new technologies to sequence DNA that were automated and less expensive.

Basically, to sequence DNA, you place all of the enzymes and nucleotides (A, G, C and T) necessary to copy DNA into a test tube. A small percentage of the nucleotides have a fluorescent dye attached to them (a different color for each type). You then place the DNA that you want to sequence into the test tube and let it incubate for a while.

During the incubation process, the sample DNA gets copied over and over again. For any given copy, the copying process stops when a fluorescent nucleotide gets placed into it. So, at the end of the incubation process, you have many fragments of the original DNA of varying sizes and ending in one of the fluorescent nucleotides. For an animation of this process of DNA sequencing, visit [DNA Interactive](http://www.dnai.org/b/index.html), go to Techniques, then Sorting and sequencing.

DNA technology will continue to develop as we try to understand how the elements of the human genome work and interact with the environment.

For lots more information about DNA and related topics, check out the links on the site.

**PHENYLKETONURIA**

A gene in your DNA codes for an **enzyme** (type of protein that speeds up a chemical reaction) that allows you to break down a specific amino acid called **phenylalanine** found in milk. In some individuals, this gene is missing or defective. They do not make the enzyme and cannot break down phenylalanine. As infants, if these individuals receive regular milk or milk products, the unbroken phenylalanine will accumulate and cause brain damage (The disease is called**phenylketonuria**). Fortunately, a genetic test is conducted at birth that can identify these babies, who can then be fed milk products lacking phenylalanine (such as soy milk).