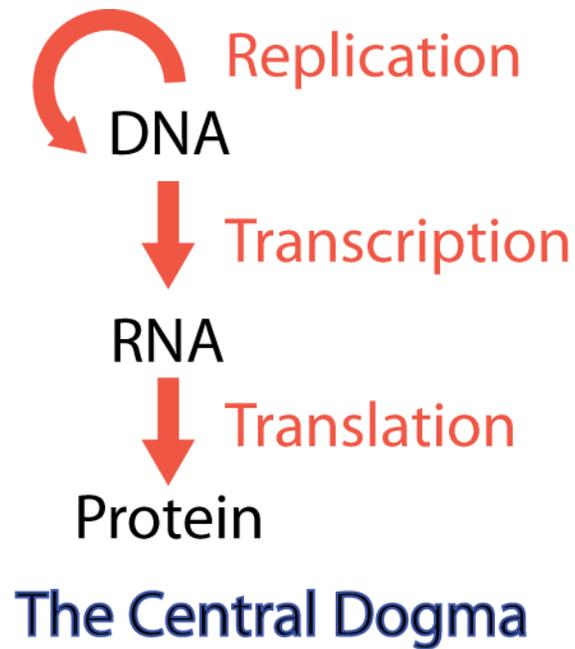


Basic Molecular Biology

The Central Dogma

Alright, we've described the basic players in molecular biology: DNA, RNA, and proteins. What they do is this:



Here's what's going on: You have the Genome which is a really big double-stranded DNA molecule within every cell. Every cell has a genomic DNA. Through the process of Replication, the genomic DNA can be copied into another identical molecule. When cells divide, each daughter cell gets a copy of the genomic DNA. Short regions of the genome called genes (typically 1/1000th of the genome's total length) get Transcribed into RNA molecules. RNAs then get Translated into protein molecules. Proteins do all the busy work.

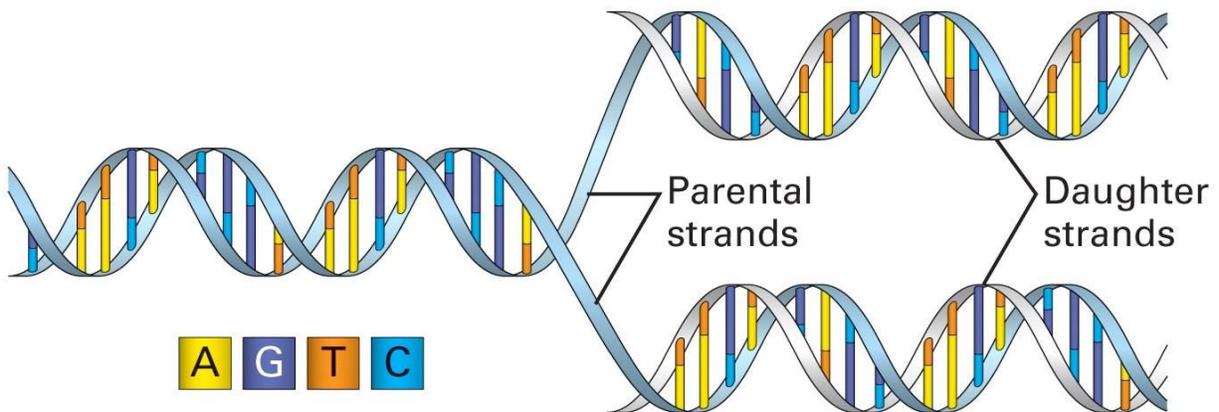
The thing that is being preserved during each of these steps is information. The sequence of DNA is transcribed base-per-base using simple rules into RNA. The RNA is translated according to a simple code called the genetic code 3 bases at a time into a sequence of amino acids. Since the number of different genes present in a cell is usually between 500 and 50,000, many different proteins are encoded by a cell. Those proteins go on to be enzymes that catalyze chemical reactions, be structural materials, transport chemicals, and do all sorts of other amazing things. So, how does it work? Well, it's very complicated, and I will only give you the 30 thousand-foot view of what's going on. These processes of replication, transcription, and translation are themselves performed by proteins, RNA molecules, and a few small molecules like the amino acids and nucleotides. So, one of the roles of the proteins encoded by the genome is to make the machinery that does the central dogma. Additionally,

the smorgasbord of chemical reactions that make up primary metabolism are each performed by at least one protein encoded by a gene in the genome.

To really understand what's going on in the Central Dogma, any basic modern biology, biochemistry, or molecular biology textbook will go through it in great detail. Additionally, there are all sorts of web pages that go through it. I'll just touch on the basics:

Replication

The core molecular machine responsible for replication is the "DNA polymerase". There are many proteins that together make up the DNA polymerase complexes, but in general they start their work on a specific sequence of DNA on the genome called the "origin". What the polymerases do is first break apart the two strands of DNA and use deoxynucleotide monomers to polymerize a new strand of DNA complementary to the first. The result of replication is two double-stranded DNAs that are identical to the original.



Transcription

Additional DNA elements called "promoters" and "terminators" define the boundaries of genes. RNA polymerases are the molecular machines that look for promoter elements in the DNA and initiate the polymerization of ribonucleotides into RNAs complementary to the DNA. Transcription proceeds much like replication. It starts at the promoter (rather than the origin), opens up the double stranded DNA and starts adding NTPs to a growing chain. It stops when it reaches the terminator and releases the free single-stranded RNA molecule which is now called a "messenger RNA" or "mRNA".

Translation

Translation acts up the mRNA molecule product of transcription. It involves amino acids, a large molecular machine composed of protein and RNA molecules called the ribosome, RNA molecules called transfer RNAs or "tRNA", and assorted other proteins. In general, the ribosome searches the RNA molecule for a specific signal sequence called the ribosome binding site and then initiates polymerization of a peptide chain at the "start codon" which is usually an AUG. It then adds amino acids

one at a time by reading the next 3 bases and adding the amino acid corresponding to that 3 base sequence (called a “codon”) according to the genetic code:

The Genetic Code

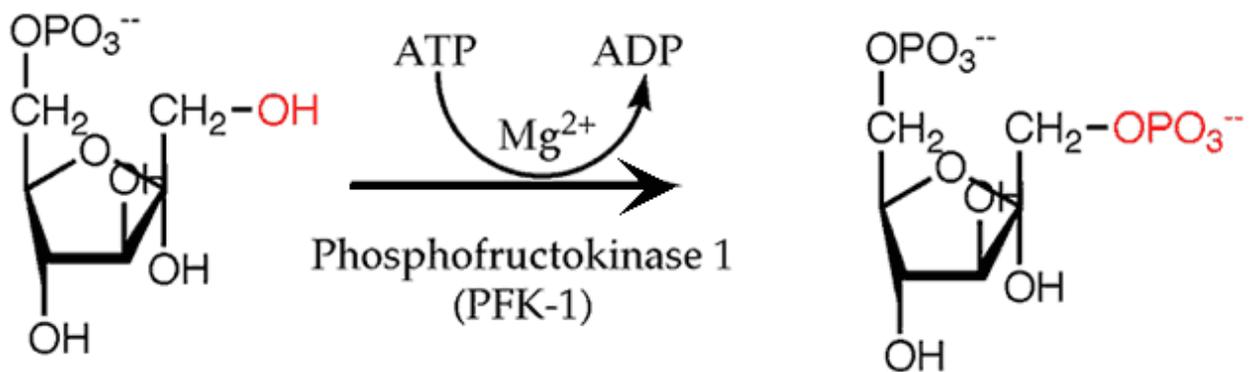
	U	C	A	G	
U	UUU Phenylalanine UUC alanine UUG Leucine UUA Leucine	UCU Serine UCC Serine UCA Serine UCG Serine	UAU Tyrosine UAC Tyrosine UAA Stop UAG Stop	UGU Cysteine UGC Cysteine UGA Stop UGG Tryptophan	U C A G
C	CUU Leucine CUC Leucine CUA Leucine CUG Leucine	CCU Proline CCC Proline CCA Proline CCG Proline	CAU Histidine CAC Histidine CAA Glutamine CAG Glutamine	CGU Arginine CGC Arginine CGA Arginine CGG Arginine	U C A G
A	AUU Isoleucine AUC Isoleucine AUA Isoleucine AUG Methionine	ACU Threonine ACC Threonine ACA Threonine ACG Threonine	AAU Asparagine AAC Asparagine AAA Lysine AAG Lysine	AGU Serine AGC Serine AGA Arginine AGG Arginine	U C A G
G	GUU Valine GUC Valine GUA Valine GUG Valine	GCU Alanine GCC Alanine GCA Alanine GCG Alanine	GAU Aspartic acid GAC Aspartic acid GAA Glutamic acid GAG Glutamic acid	GGU Glycine GGC Glycine GGA Glycine GGG Glycine	U C A G

The tRNA molecules are the adapters that “read” the RNA and match up each codon with an amino acid. These tRNA molecules have a region called the “anticodon” that base-pairs with the mRNA codons. On one end of the tRNA is a covalently attached amino acid. When the ribosome finds a match between the tRNA and the current codon, it will transfer the covalently-charged amino acid onto the growing chain. The ribosome keeps doing this until it encounters one of the 3 stop codons and then it releases the newly-synthesized protein. That new protein can then fold into its functional form and do whatever biochemical function it can.

Regulation

The biochemical composition and behavior of a cell is somehow all orchestrated by its genes and biochemical products. To make it all work, the genome is encoding tons of information about the who, what, where, when and how much of the biochemistry of the cell. The composition of the cell—the who and what—of a cell is determined by the protein sequences encoded by the genes and the biological functions those proteins exert on the cell. So, for example, myosin is the major protein in muscle that has the biological function of converting chemical energy into physical movement. Myosin is transcribed 3 bases at a time into the amino acid sequence that has this biochemical activity. That information is readily apparent in the gene sequence as the rules for encoding DNA sequence into protein is pretty simple. So, the “who” of molecular movement is the Myosin protein molecule, and the “what it does”

of molecular movement is a chemical property of the protein molecule itself. The “what it does” is the real magic of biology. It is determined by the amino acid sequence of a protein, but how you go from basic chemical principles to understanding the biochemical behavior of a protein is really complicated. Nevertheless, it is sufficient to appreciate that “what it does” is somehow a function of the amino acid sequence of itself—the molecule knows what it’s supposed to do. Take another example: phosphofructokinase. This is an enzyme that catalyzes the addition of a phosphate moiety from ATP to fructose-6-phosphate. It is one of many protein enzymes involved in the process of extracting the chemical energy from glucose and transferring it to ATP currency in the cell. “Who it is” is a molecule of defined amino acid sequence encoded by its gene. “What it does” is catalyze a chemical reaction involving ATP and fructose-6-phosphate. The ability to do this reaction is an intrinsic property of the protein molecule—no additional information is needed to make the protein molecule do its reaction.



Regulation deals with the where, when, and how much of a biochemical in a cell. There are many distinct environments within a cell. A bacterium has its cytoplasm, membrane, and periplasm. A human cell has organelles such as mitochondria, the endoplasmic reticulum, lysosomes, the nucleus, etc. Cells also are able to send molecules outside of the cell. So, the “where” of a biochemical somehow has to be dictated through regulation. There has to be some sort of information provided in the gene that tells the cell where that molecule is supposed to go. I won’t explain how it works, but suffice it to say that you often have short regions of a protein, often on the ends, that encode the information about where a protein goes. Small molecules tend to exist wherever the enzymes that make them exist.

The “How much” and “when” is where regulation gets complicated. This is where the genetic circuits and regulatory networks come into play. There are many many many mechanisms that cells use to do this type of regulation. You can distinguish the types of control by where they are exerted within the central dogma. For example, many genes are regulated at the level of transcription. By this, we mean that the gene isn’t always being described or it might be transcribed at different levels depending on the internal or external environment of a cell. For example let’s say there isn’t any glucose present in the environment of a bacterium. The bacterium has no glucose to consume, so why would it bother making the enzymes needed to breakdown glucose? So, you might expect that the gene for phosphofructokinase might be “repressed” in the absence of exogenous glucose. Exactly how that

works gets complicated, but in general, somehow the RNA polymerase within the cell only works on the promoter of the phosphofructokinase gene when external glucose is present.

You can also have translational control, though it is not generally as common a mechanism as transcriptional control. Here, some biochemical in the cell is acting upon the mRNA of a gene and preventing the ribosome from translating it. There can be all sorts of posttranslational control which is extremely common. Here, biochemicals in the cell act on the protein through protein-protein interactions, small molecule-protein interactions, covalent modifications such as phosphorylation, etc. to alter the proteins activity or even cause the protein to degrade. Finally, you can have intrinsic regulation that is the property of the protein itself. Product inhibition is one of these mechanisms. Here, an enzyme for a chemical reaction is inhibited by the product of the reaction it catalyzes. The result is that it only makes so much product and then shuts itself down—a simple negative feedback mechanism. This type of behavior is very common and robust, though, particularly in primary metabolism.

So, regulation is really complicated, and I talk about it here because 1) it's an incredibly important part of how cells work and interact with each other and their environment, and 2) it's primarily what most synthetic biology folks are trying to engineer.

How do I learn more?

This pretty much covers the basic principles of molecular biology and biochemistry. Understanding the rest of it is largely a matter of learning all the various details. I highly recommend the textbook “Molecular Biology of the Cell”. One thing to keep in mind is that there are 3 distinct domains of life: prokaryotes (the bacteria), eukaryotes (plants, animals, fungi), and the archaea (funky single-celled things that live in places like hydrothermal vents). The rules for molecular biology are a little different for the 3 different domains. Pretty much everything I've said here is universal, but as you build up on this foundation things start diverging. So, as you are reading something, keep in your mind what domain they are talking about and don't mix up the stories in your mind.