BIOL 300 – Foundations of Biology
Summer 2017 – Telleen
Lecture Outline

RNA, the Genetic Code, Proteins

I. How RNA differs from DNA

A. The sugar ribose replaces deoxyribose. The presence of the oxygen on the 2' carbon atom of the sugar (-OH instead of –H) makes RNA much more chemically reactive than DNA

B. The base uracil (U) replaces the base thymine (T). Although chemically similar, the replacement of T with U marks the molecule as RNA. U still base pairs with A.

C. RNA is usually single-stranded, but can loop back on itself to form double-stranded regions and what is called secondary structure

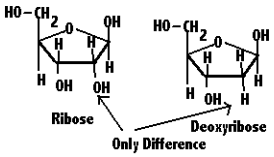
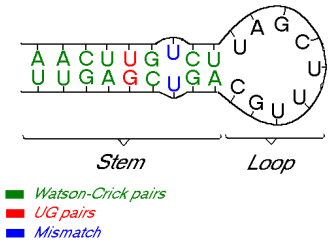
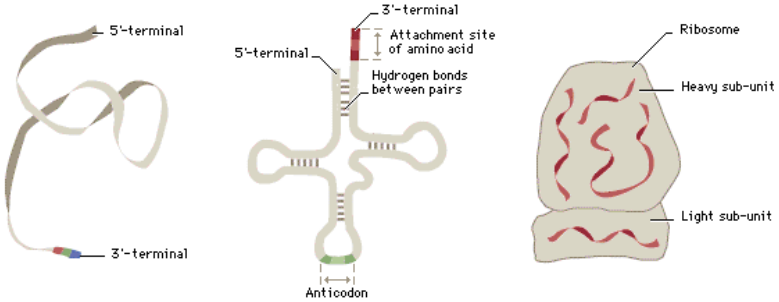


Fig 2- Ribose and Deoxyribose



D. There are three major classes of RNA molecules:

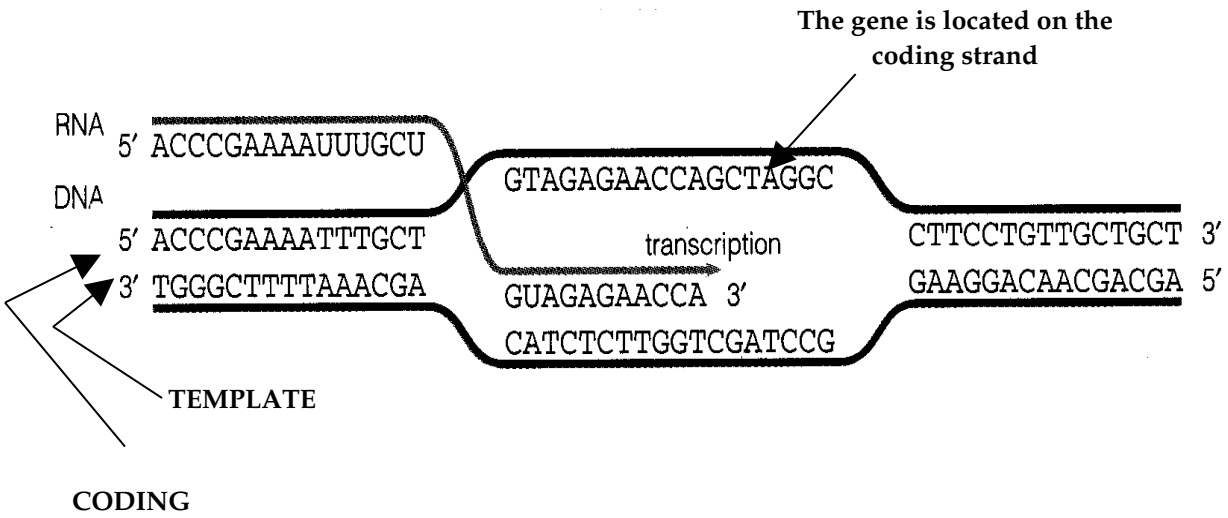
1. mRNA – messenger RNAs carry information for the building of polypeptides. They are transcribed from genes and translated into polypeptides
2. rRNA – ribosomal RNAs are small RNAs that do not carry information. They are transcribed from genes, and are involved in translations but are not themselves translated
3. tRNA – transfer RNAs are are small RNAs that do not carry information but help in translation by carrying amino acids
4. Other small RNAs include miRNA, siRNA, and snRNA



II. How genes are read

A. Each gene is read and copied into RNA by a protein complex called RNA polymerase

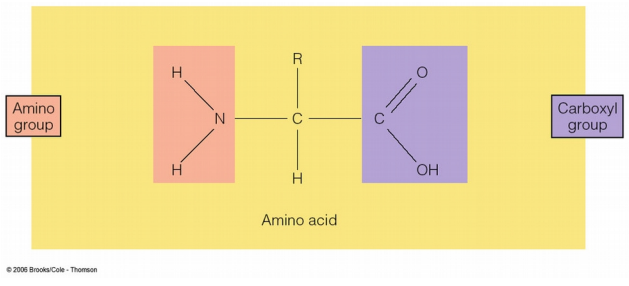
B. Steps in the process:
    1. Initiation: RNA polymerase binds to the DNA just in front of a gene called the promoter where it opens up the DNA (separates the strands) to form a transcription bubble
    2. RNA polymerase begins copying one strand of the DNA into a single-stranded RNA molecule. The RNA pol. reads the template strand, which results in an RNA version of the coding strand. This process is called elongation
    3. Sequences at the end of a gene, called the terminator, tell the RNA pol. to stop
    4. The newly made RNA them drop off the DNA and the DNA comes back together



III. Amino Acids, Polypeptides, and Proteins

A. A protein is one of a group of high molecular weight, nitrogen-containing organic molecules of complex shape and composition
    1. Each cell type has a characteristic set of proteins that give the cell its functional properties
    2. A protein is a functional unit and can be comprised of more than one polypeptide (a polymer of amino acids linked together)
    3. Each polypeptide has a specific number, kind, and sequence of amino acids that give it its 3-D shape and biochemical properties

B. Life uses 20 different left-handed amino acids
    1. Each one has different chemical properties, but they can be put into groups with similar properties (eg. Acidic, basic, hydrophilic, hydrophobic, etc)



IV. The Genetic Code of Life

A. The amino acid sequence of a polypeptide is coded by the sequence of nucleotides in a gene.

1. There are 20 amino acids and only 4 nucleotides, so it can't be a single letter code.
2. It can't be a doublet (or two-letter) code because there would still only be $4^2 = 16$ possibilities which is too few to code for the 20 amino acids we observe
3. The genetic code is a triplet (three-letter) code. Since $4^3 = 64$ possibilities, there are more than enough three letter combinations (codons) to account for the 20 amino acids. Many amino acids are coded for by multiple codons, so we say that the genetic code is degenerate. There are also STOP codons that halt translation

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

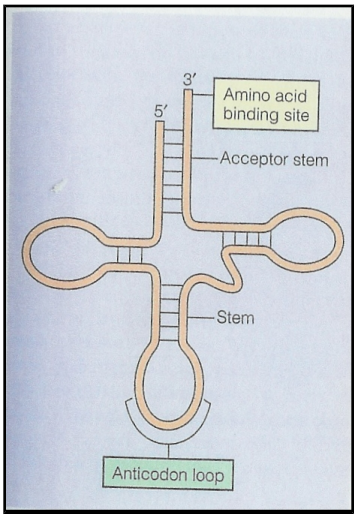
V. The role of rRNAs in translation

A. Proteins synthesis takes place on huge enzyme complexes called ribosomes

1. Ribosomes are in the cytoplasm of the cell
2. Ribosomes consist of both Protein and rRNA
3. During translation, the ribosome holds together the mRNA being translated, the newly forming polypeptide, and tRNA molecules (which we'll discuss in a minute)

VI. The role of tRNAs in translation

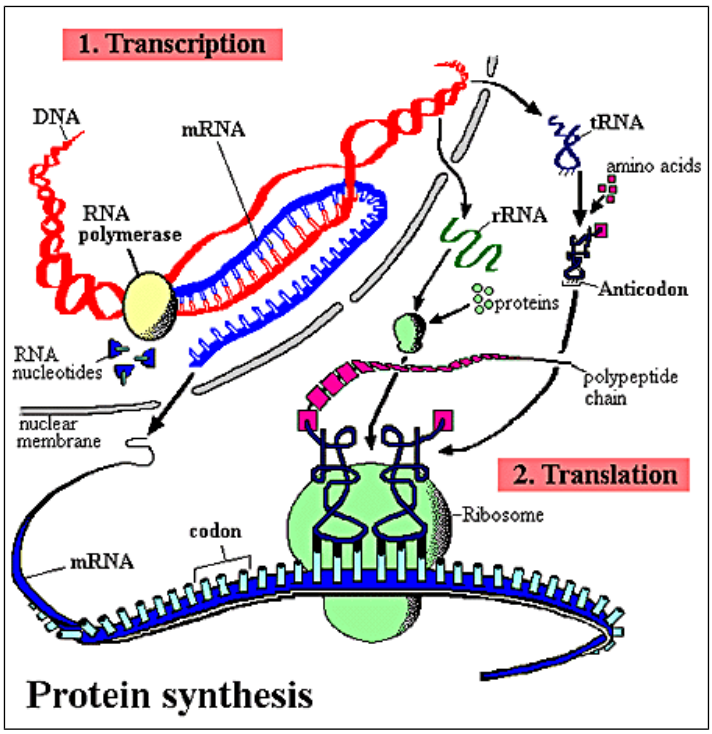
A. tRNA molecules bring the amino acids to the ribosome



1. They are the actual translator molecule that relates a codon to an amino acid
2. They "know" the chemistry of both nucleic acids and amino acids
3. tRNAs have two ends: the amino acid-carrying end and the anticodon end
4. The amino acid end binds to an amino acid
5. The anticodon end (which is part of a stem-loop structure) base pairs with the codon in the mRNA
    a. If the codon is 5'-AUG-3', the corresponding anticodon would be 5'-CAU-3' and the tRNA would carry Methionine since that is the amino acid that AUG codes for

VII. Putting it all together: The process of translation

A. After transcription, the mRNA leaves the nucleus and enters the cytoplasm
B. In the cytoplasm, the mRNA is immediately bound by a small ribosomal subunit
C. The first codon in the mRNA that is read by the ribosome is always AUG, which codes for methionine. This is called the start codon
D. A tRNA with anticodon 3'-UAC-5' (met-tRNA) joins the complex by base pairing with the codon
E. A large ribosomal subunit joins the complex. This creates two enzymatic sites called the P (peptidyl) site and the A (acceptor) site. The met-tRNA is in the P site, while the A site is empty.
F. A tRNA with an anticodon matching the second codon in the mRNA enters the empty A site and base pairs with the second codon.
G. The bond between methionine and its tRNA is broken and it is transferred onto the amino acid carried by the tRNA is the A site. Note that the amino acid in the A site now carries two amino acids. The two amino acids have now formed a 2 unit peptide chain.
H. The ribosome shifts three bases along the mRNA in a process called translocation
I. The met-tRNA is released and the second tRNA is now in the P site, while the A site is empty.
J. The tRNA corresponding to the next codon comes in and the process repeats until a stop codon is reached (there are no tRNAs for stop codons)
K. Once a stop codon is reached, the complex releases to mRNA and the polypeptide and disassociates.



Gene Structure, Regulation of Gene Expression, and Mutations at the DNA Level

I. Gene Structure

A. We already talked briefly about gene structure in general terms, but now we are going to expand it.

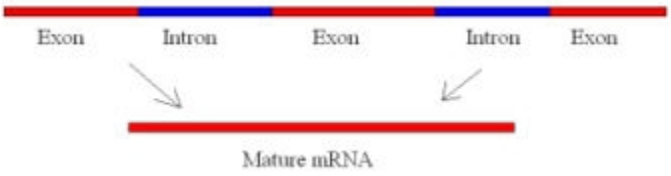
B. Our early model of gene structure just contained a promoter, a protein coding DNA sequence, and a terminator, but that is an oversimplification.

C. Scientists often use the term **Open Reading Frame, or ORF**, to describe the protein coding sequence of a gene. An **ORF** is a **start codon** (ATG) followed by a long stretch of amino acid coding sequences and then a **stop codon**.

D. However, this is complicated by the fact that not all ORFs are continuous in a DNA sequence.

E. **Introns** and **Exons**

1. The DNA sequence of a gene that is transcribed into mRNA often contains extra sequence that is removed from the RNA molecule prior to translation.
2. These removed sequences are called introns
3. The sequences that remain and are translated are called exons
4. Introns are spliced out of the mRNA molecules before they are exported from the nucleus
5. Introns are sliced out based on specific sequences on the 5' and 3' ends.



II. Regulation of Gene Expression

A. From Genotype to Phenotype

1. As we have seen, genes contribute to phenotype, but the correlation is not always as straightforward as we might expect.
2. In some cases, alternative alleles are just mutations that alter or abolish the function of the protein the gene codes for. This could be introducing a premature stop codon, changing an amino acid, or altering an intron splice site.
3. Sometimes alternative alleles don't differ in the protein coding sequence, but instead change how the gene is regulated.
4. Some proteins need to be expressed in certain cells and not in others. Alternative alleles in this case could cause the expression of a particular gene in places it isn't normally (resulting in a different phenotype), or result in the protein not being present when it normally is.
5. Others result in changes in the amount of mRNA made
6. Now let's examine these cases in a little more detail.

B. Enhancers and Repressors

1. The **promoter** is responsible for initiation of transcription of a gene, but other sequences contribute to this process.
2. **Enhancers** are sequences that increase the transcription of a gene under certain conditions
3. **Repressors** are sequences that shut off the transcription of a gene under certain conditions
4. Enhancers and repressors can be located in various places relative to the ORF, including near the promoter, near the terminator, and in introns.
5. These sequences are often the binding sites for **transcription factors**, which are proteins whose role is to modulate the transcription of other genes.
6. For example, imagine a gene responsible for pigment production in f lowers (that makes anthrocyanin, a red pigment). There are transcription factors that are present in the flowers (but not the vegetative tissues) which bind to sequences near this gene and activate it. The result would be that the gene is expressed in flowers but not in leaves (because the corresponding transcription factors were not present).
7. This type of regulation is common and can get very complex when multiple enhancer and repressor sequences are involved.

C. Expression Levels

1. In addition to control of spatial and temporal control, genes are also expressed at different levels (e.g. more mRNA is made from some genes compared to others even if they are expressed in the same places and times)
2. For example, a mutation might affect how much mRNA is made from a particular gene. The normal allele might make 10,000 copies of the mRNA per cell, but an alternative allele might reduce this number to only 1,000 copies. In this case, the alternative allele may not produce as much of the protein and have a different phenotype (even though the protein itself is the same).
3. This is controlled by promoter sequences, as well as enhancers and repressors.

III. Types of mutations at the DNA level

A. Mutations can be caused in several ways

1. Incorrect copying during DNA replication
2. Environmental factors, such as chemicals and radiation (including UV light from the Sun!)

B. Mutations can sometimes be corrected by DNA repair machinery

C. The simplest types of mutations involve single nucleotide changes in the ORF of a gene.

1. Nonsense mutations are mutations that result in a premature stop codon (e.g. ATG GGA CCT **TAC** … to ATG GGA CCT **TAA** …)

2. **Missense mutations** are mutations that change an amino acid but do not result in a premature stop
   (e.g. ATG GGA CCT **TAC** … to ATG GGA CCT **TCC**…)
3. **Silent mutations** are point mutations that do not change any amino acids (e.g. ATG GGA CCT **TAC**… to ATG GGA CCT **TAT**…)

D. DNA sequences can also be mutated by insertions of extra nucleotides or deletions of nucleotides.
1. **Insertions** and **deletions** can alter an ORF as illustrated in this example
2. Imagine a series of codons reads like the following sentence:
   THE FAT CAT ATE HIS HAT
3. Now suppose an extra A is inserted into the middle of the sequence:
   THE FAA TCA TAT EHI SHA T
4. A deletion might have a similar effect:
   THE FTC ATA TEH ISH AT
5. In both cases the new sentence doesn't make sense after the insertion or deletion, showing how a protein coding sequence can be disrupted
6. Note that insertions or deletions in multiples of 3 would not have the same effect because they do not change the surrounding codons the same way. (e.g. THE FAT CAT ATE HIS NEW HAT)

E. Changes in the DNA sequences outside of the ORF can also cause phenotypic changes due to alterations of gene regulation
1. Mutations in the promoter sequence can stop genes from being expressed or alter the levels of expression
2. Mutations in enhancers/repressors can cause genes to be expressed at inappropriate times and places or alter the levels of expression
3. Mutations that affect the splice sites in introns can result in aberrant protein even if the mutation wasn't in the ORF itself.