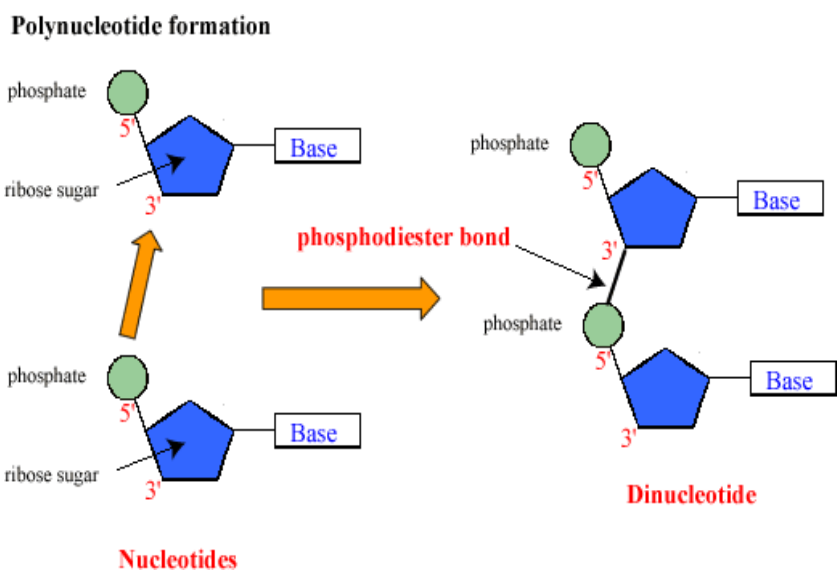
Class Outline

1. Understanding polynucleotide structure (Read)
2. Why polynucleotides are important (Read)
- Central dogma- transcription and translation
- HIV and Reverse transcription
3. Antiretroviral drug development
4. Structural biology e.g. NMR spectroscopy (also X-ray crystallography)

I have compiled these Wikipedia excerpts for you to read in preparation to the class. Reading these at home will allow us more time to talk about even more interesting things while in class. I will post any other information that I think will be useful. Please email me at mmpossi@stanford.edu with any questions.

POLYNUCLEOTIDE STRUCTURE

Deoxyribonucleic acid (**DNA**) is a molecule that contains the genetic instructions used in the development and functioning of all known living organisms and many viruses. DNA along with RNA and proteins, is one of the three major macromolecules(large complex molecules) essential for all known forms of life. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine) recorded using the letters G, A, T, and C. Most DNA molecules are double-stranded helices, consisting of two long polymers of simple units called nucleotides, molecules with backbones made of alternating sugars (deoxyribose) and phosphate groups (related to phosphoric acid), with the nucleobases (G, A, T, C) attached to the sugars. DNA is well-suited for biological information storage, since the DNA backbone is resistant to cleavage and the double-stranded structure provides the molecule with a built-in duplicate of the encoded information.

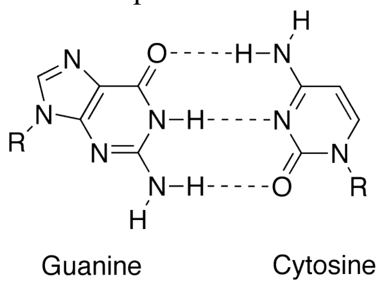


This figure shows how the backbone is formed-the arrangement of the sugar, phosphate and base in a DNA molecule

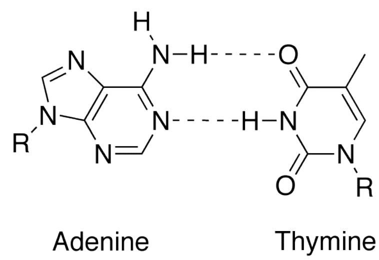
These two strands run in opposite directions to each other and are therefore anti-parallel, one backbone being 3′ (three prime) and the other 5′ (five prime). This refers to the direction the 3rd and 5th carbon on the sugar molecule is facing. Attached to each sugar is one of four types of molecules called nucleobases (informally, *bases*). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription.

Within cells, DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts. In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

G-C base pair



A-T base pair



A DNA double helix is made of various combinations of these ***base pairs****. G always paired with C and A to T.* ***R*** *represents the phosphate group and ribose sugar attached to the base. We will talk more about these structures in class.*

Ribonucleic Acid (RNA). In comparison with DNA:

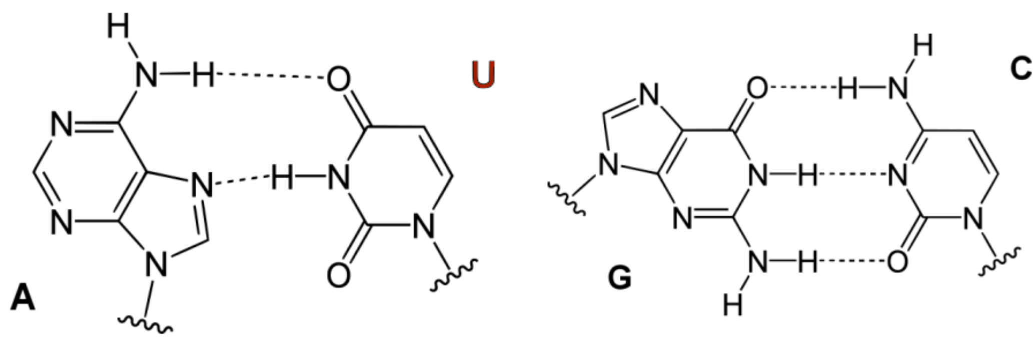
The chemical structure of RNA is very similar to that of DNA, but differ in three main ways:

- Unlike double-stranded DNA, RNA is a single-stranded molecule in many of its biological roles and has a much shorter chain of nucleotides. However, RNA can, by complementary base pairing, form intra-strand double helixes, as in tRNA through base pairing.
- While DNA contains *deoxyribose*, RNA contains *ribose* (in deoxyribose there is no hydroxyl group attached to the pentose ring in the 2' position). These hydroxyl groups make RNA **less stable** than DNA because it is more prone to hydrolysis.
- The complementary base to adenine is not thymine, as it is in DNA, but rather uracil, which is an unmethylated form of thymine.[1]

Like DNA, most biologically active RNAs, including mRNA, tRNA, rRNA, snRNAs, and other non-coding RNAs, contain self-complementary sequences that allow parts of the RNA to fold and pair with itself to form double helices. Analysis of these RNAs has revealed that they are highly structured. Unlike DNA, their structures do not consist of long double helices but rather collections of short helices packed together into structures akin to proteins. In this fashion, RNAs can achieve chemical catalysis, like enzymes. For instance, determination of the structure of the ribosome—an enzyme that catalyzes peptide bond formation—revealed that its active site is composed entirely of RNA.

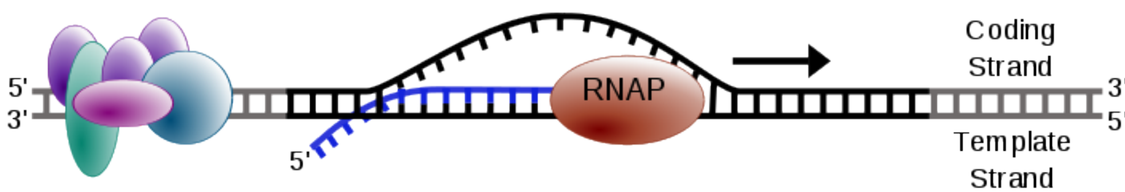
RNA base pairs

A-U base pair **G-C base pair**



HOW DNAs AND RNAs ARE USED BY LIVING CELLS

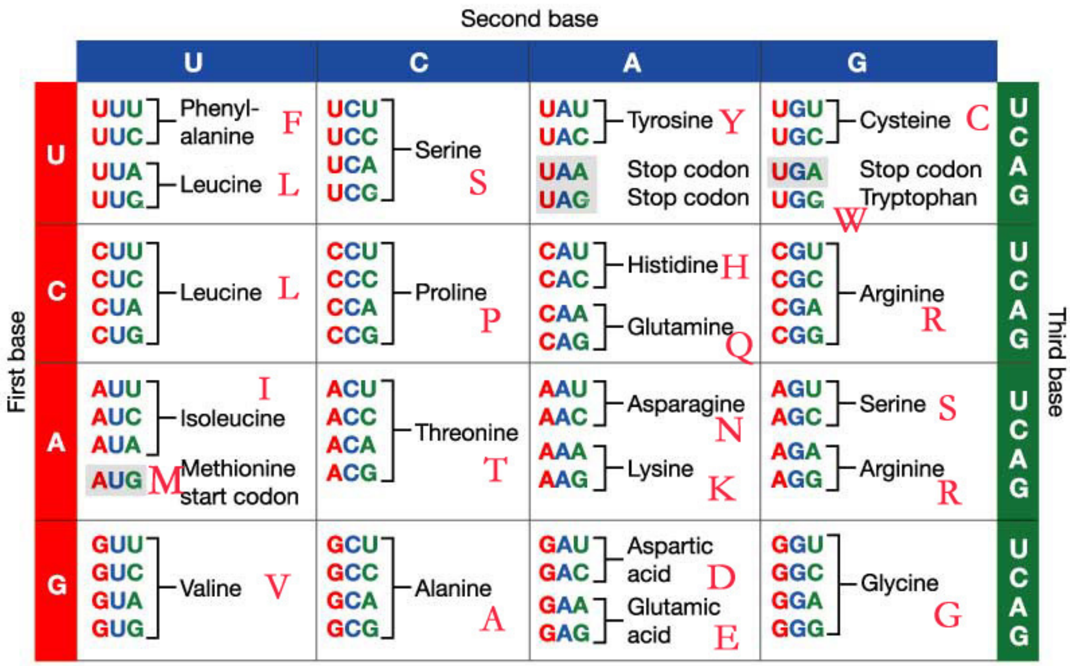
Transcription is the process by which the information contained in a section of DNA is transferred to a newly assembled piece of messenger RNA (mRNA) and other types of RNA (e.g. ribosomal RNA, transfer RNA etc.). It is facilitated by RNA polymerase and transcription factors (a bunch of proteins which bind DNA). In eukaryotic cells the primary transcript (pre-mRNA) must be processed further in order to ensure translation (e.g. cutting parts of the primary RNA known as RNA splicing). Alternative splicing (different ways of cutting RNA strands) can also occur, which contributes to the diversity of proteins any single mRNA can produce. mRNA that do this are known as polycistronic RNA.



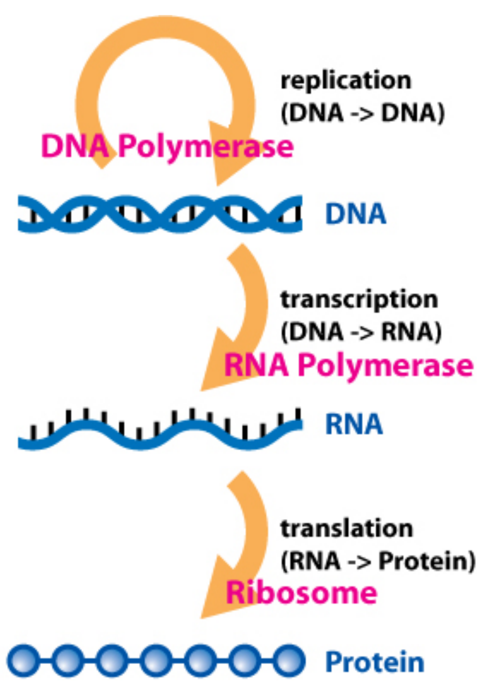
RNA polymerase (RNAP) transcribing DNA from the 5' end to the 3' end.

(Have a general understanding in this next part i.e. translation- don't pay attention to small details. You should know, among other things, what tRNAs do in translation)

Translation. Eventually, this mature mRNA finds its way to a ribosome, where it is translated. In eukaryotic cells, the site of transcription (the cell nucleus) is usually separated from the site of translation (the cytoplasm), so the mRNA must be transported out of the nucleus into the cytoplasm, where it can be bound by ribosomes. The mRNA is read by the ribosome as triplet codons, usually beginning with an AUG (adenine−uracil−guanine), or initiator methionine codon downstream of the ribosome binding site. Complexes of initiation factors and elongation factors bring aminoacylated transfer RNAs (tRNAs) into the ribosome-mRNA complex, matching the codon in the mRNA to the anti-codon on the tRNA, thereby adding the correct amino acid in the sequence encoding the gene. As the amino acids are linked into the growing peptide chain, they begin folding into the correct conformation. Translation ends with a UAA, UGA, or UAG stop codon. The nascent polypeptide chain is then released from the ribosome as a mature protein. In some cases the new polypeptide chain requires additional processing, like folding and/or splicing, to make a mature protein.



A codon-amino acid chart



A summary of the path taken by DNA to achieve effect as protein or as RNA

DNA replication. Before proceeding to the next section please read this page.
http://en.wikipedia.org/wiki/DNA_replication
Among other things, you should understand that DNA polymerase catalyses polymerisation reactions by elongating a complementary RNA primer, usually constructed on DNA template by the enzyme DNA primase. In the case of reverse transcription discussed below, tRNAs are used as a primer!

HIV and Reverse Transcription.

I will talk about **HIV** in class, so if you haven't heard much about the virus- it is truly fascinating and I would look up the Wikipedia page here.

Reverse transcription is a transcription reaction in retroviruses (e.g. Human Immunodeficiency Virus (HIV)) used to make a DNA molecule from an RNA template using the enzyme reverse transcriptase (RT). This process occurs in the cytosol of the cell and has the following features:

1. A specific cellular tRNA acts as a primer and hybridizes to a complementary part of the virus genome called the primer binding site (PBS)
2. Complementary DNA then binds to the U5 (non-coding region) and R region (a direct repeat found at both ends of the RNA molecule) of the viral RNA
3. A domain on the reverse transcriptase enzyme called RNAse H degrades the 5' end of the RNA which removes the U5 and R region
4. The primer then 'jumps' to the 3' end of the viral genome and the newly synthesized DNA strands hybridizes to the complementary R region on the RNA
5. The first strand of complementary DNA (cDNA) is extended and the majority of viral RNA is degraded by RNAse H
6. Once the strand is completed, second strand synthesis is initiated from the viral RNA
7. There is then another 'jump' where the PBS from the second strand hybridizes with the complementary PBS on the first strand
8. Both strands are extended further and can be incorporated into the hosts genome by the enzyme integrase

We will talk a little more about the initiation of reverse transcription in class. We will also consider tools we can use to understand this small part of the process as well as why it is important. See you in class on Sunday!

References:
Wikipedia pages on the following topics
1. DNA
2. Central Dogma
3. Translation
4. Transcription
5. Reverse transcriptase

6. RNA
7. DNA replication
8. Codon chart from BioGerm.org http://www.biogem.org/blog/rna-to-protein-translation-in-perl/