BIOMOLECULES.2 (nucleic acids, genetic code)

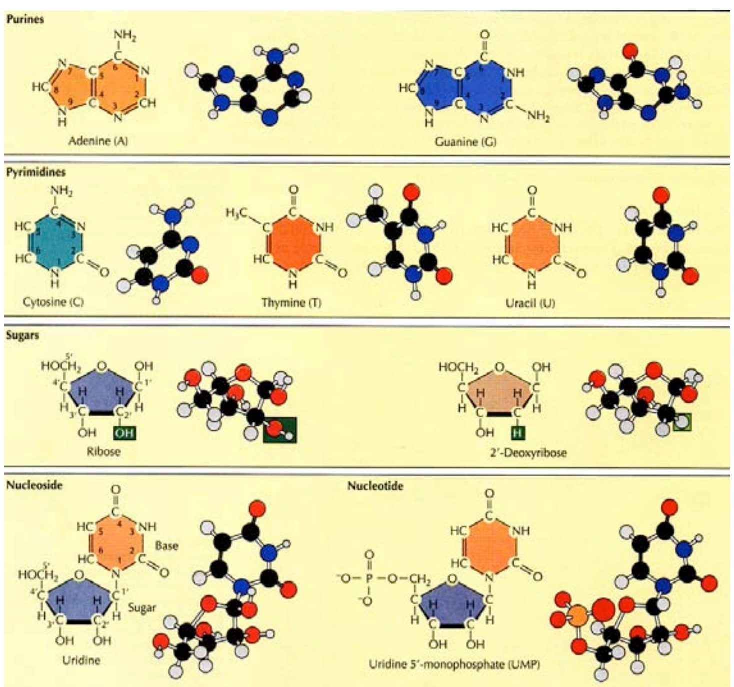
Nucleic acids -- these molecules are the basis for the genetic material of all life on Earth, and so are central for our speculations about life elsewhere. They consist of sequences of *nucleotides*, which are three chemical groups bonded together: one of four (or five) bases, a particular sugar, and a phosphate group. These nucleotides somehow became capable of linking up, or polymerizing, to form long sequences called nucleic acid, either single-stranded (RNA) or double-stranded (DNA). Before discussing these in more detail, it is easiest to study the pictures below.

1. **a *nitrogenous* base** – these 5 or 6 sided ring molecules are the "alphabet" of the genetic code. There are two classes, called purines (A, G; structurally they are two connected rings) and pyrimidines (C, T in DNA, C, U in RNA; structurally a single ring). These bases are illustrated below.

Why these four (or five) bases?? Good question!

2. a 5-carbon sugar: (ribose for RNA, 2' -Deoxyribose for DNA – just ribose with one of the O atoms removed at the "2'" site of the sugar ring)

The joined base + sugar is called a *nucleoside*. Study the picture below:



3. a phosphate ~ (P + 4 O's), where P is phosphorus.
Recall that the joined base+sugar is called a *nucleoside* (above drawing); the phosphates link together the nucleosides. The base + sugar + phosphate is called a *nucleotide*. [Don't worry about keeping these straight. I will not ask you to remember the distinction.]

That extra phosphate group turns out to be very important, and many people think that phosphorus has several unique properties that make it an optimal (and maybe the only) choice for the third monomer of nucleic acids.

Even on their own, various nucleotides serve a number of cellular functions. For example cyclic AMP (cyclic adenosine monophosphate) is used in intracellular communication; ATP (adenosine triphosphate) is crucial in energy transfer, and coenzymes (nucleotides linked to a vitamin) are active in cellular metabolisms. To illustrate the importance of the phosphates, ATP generates energy by breaking and then reforming its third ("tri-") phosphate in the phosphate side group.

The illustration below shows how the 3 monomers of DNA come together to form the "double-helix"

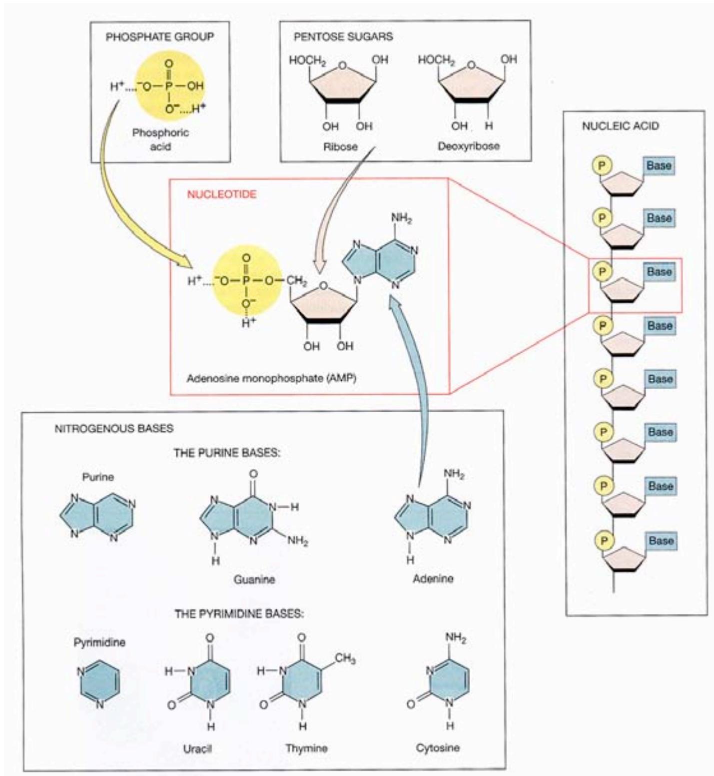


FIGURE 3–15 Nucleotides and nucleic acids. A nucleotide (center box) has three parts: one to three phosphate groups (yellow), a sugar (brown), and a base (blue-green). The ring-shaped nitrogenous bases (lower box) are derived from one of two parent compounds, purine and pyrimidine. A nucleic acid polymer (far right box) consists of nucleotide monomers, linked with the phosphate group of one attached to the sugar of the next.

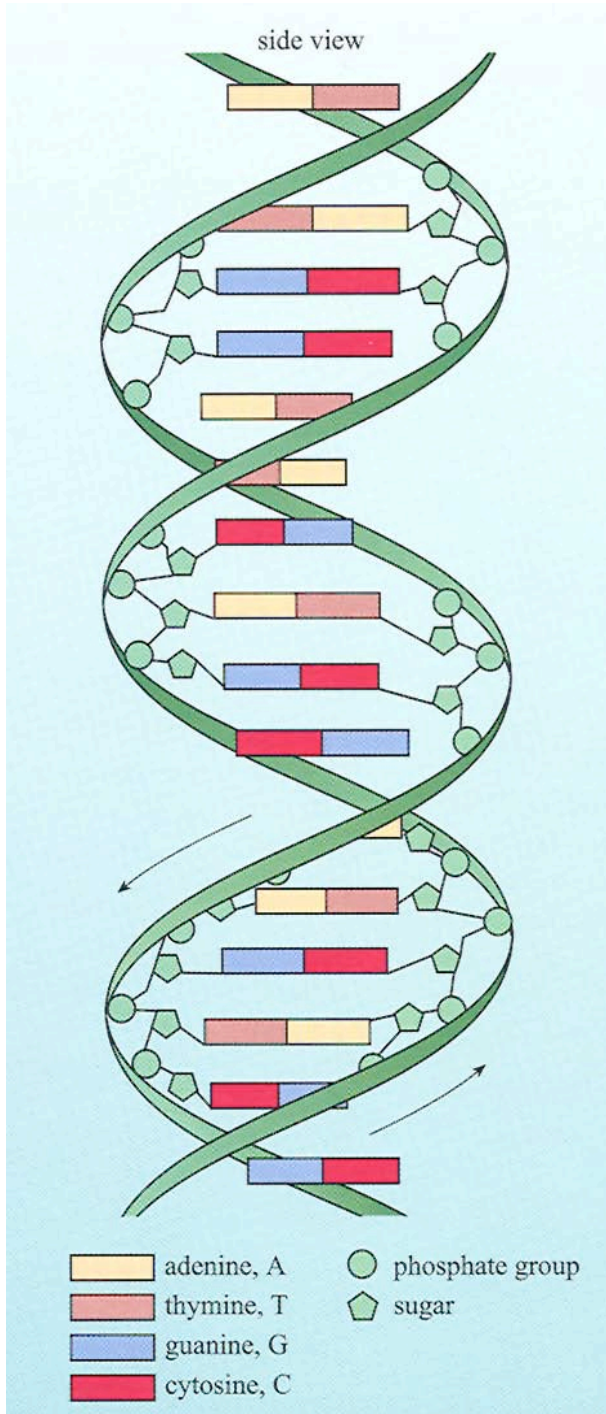
The molecules central to all terrestrial organisms and of central interest to us are long (~10^5 atoms, RNA) or *very* long (up to about 10^{10} atoms in DNA, the "double helix") *sequences of nucleotides*. These are nucleic acids here on Earth, but there could be analogous highly efficient use of sequences to store information on other worlds.

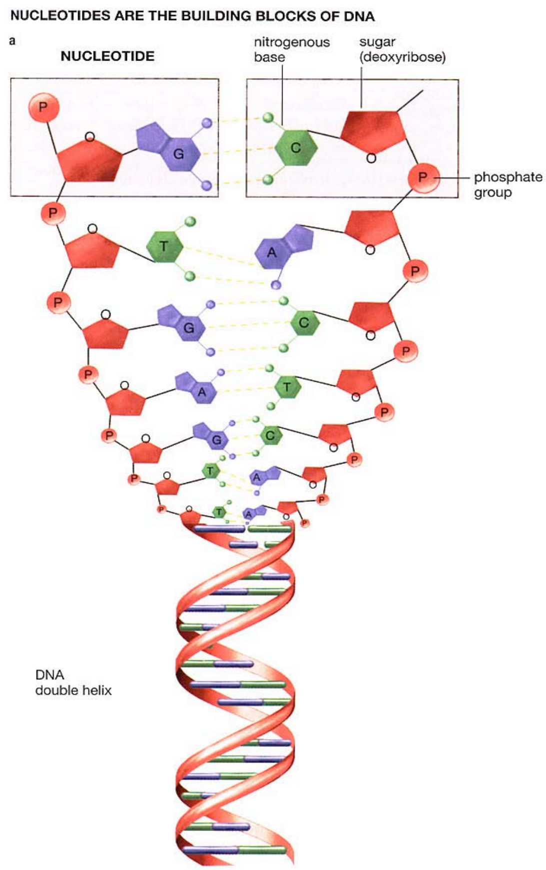
Please note that DNA is remarkably simple—it is just a chain of nucleotides. Also note that if it you could straighten out DNA, it would be about a mm (bacteria) to a cm (e.g. vertebrates) long! It might be impossible to find a compartment (cell membrane) large enough to house such a large information-carrying molecule. Instead, the double-helix coiling of the DNA allows it to be contained in a regions smaller than a *micro*meter. If DNA wasn't so coiled, there might never have been genetic material capable of carrying so much information.

Again we see that part of the remarkable properties of complex biomolecules is their ability to attain varied and/or important shapes (compare with proteins and lipids earlier).

In DNA the two strands are wound around each other, joined by ***base-pairing*** between each strand. The key feature of DNA is that each base can *only* be paired with its "complementary" base: A with G, C with T. Note that the bases are joined by hydrogen bonds (discussed earlier). Base-pairing is the key to replication in DNA.

Study the pictures below:


side view
adenine, A
thymine, T
guanine, G
cytosine, C
phosphate group
sugar



These sequences work like an alphabet, "spelling out" words ("<u>codons</u>", a sequence of three bases) that specify particular amino acids. A sequence of codons (words) that specify the amino acids needed to make a particular protein is called a *<u>gene</u>*. Because they "code" for proteins, these polymers are often called "informational macromolecules". Other parts of genes (surprisingly large fractions!) do not code for genes, and used to be called "junk DNA" or introns. Today these **noncoding sequences** are recognized as extremely complex systems of networks that regulate gene expression, i.e. when various genes function or are "silenced." Special kinds of RNA turn out to also play an important role in gene expression, but there is too much detail for us to cover.

Genetic code and replication process

translation dictionary = genetic code:

a codon of 3 bases (out of 4) is a triplet code for specifying an amino acid, e.g. in RNA (single strand):

```
P---S---P---S----P----S
          |   |   |  |   |   |
          A   C   U  <-----a  codon
```

(Note: no. of possible codons = $4^3 = 64$, which is greater than 20. Think about it!)

A gene = sequence of codons long enough to specify a protein (~100-500 triplets long)

In DNA, bases can't pair at random. Only A--T, G--C (base pairs). When the 2 DNA strands unwind, each half can reproduce its partner exactly.

Messenger RNA reads info (codons) from open DNA file. Message taken to ribosome = assembly line for construction of proteins; made of ~50 protein + RNA molecules. The needed amino acids are brought to the ribosome by various transfer RNAs.

Can think of ``life" as a protein-making gene system.

How was this accomplished? With great complication!

DNA—RNA—protein coupling in terrestrial organisms:

transcription in nucleus | "introns" spliced out (in nucleus) | translation in ribosomes*

DNA -----------> hn RNA ----------------> mRNA---------------------> protein

e.g. "TCC" → "AGG" | e.g. "AGC"→ arginine

hn RNA: heterogeneous nuclear RNA

mRNA: messenger RNA (passes through nuclear pores)

"working copy" of part of DNA; codes for one or more proteins

***Ribosomes**: These are complex structures composed of proteins and RNA that serve as a kind of "dock;" ribosomes read bases along the messenger RNA in groups of 3 (codons), which specify 1 of the 20 amino acids (or ``stop"). This is called "translation." Ribosomes "send out" transfer RNA to gather the right amino acid. So this is an extremely complex part of the replication process.

How did all this complexity come about? Undoubtedly the process wasn't so advanced in the beginning, so most people try to look at simpler self-replication schemes relying on the same principles of coding. In particular, the essential parts of the process involve the use of an *informational sequence* that can code, and a "template" molecule (RNA here) that can be read in other parts of the cell.

A major breakthrough we will discuss later is that RNA can serve as its own template, and can in fact act like a protein, carrying out functions like repairing itself through its various types of 3-dimensional shapes (conformations). Even for life that is based on DNA, it is now becoming recognized that RNA plays a much richer role than was previously thought. Perhaps it started out as the "jack-of-all-trades" molecule for primitive life, which eventually became a DNA-protein system with RNA playing the role of intermediary, and carrying out several of its own functions.

So now that we understand a little about how the genetic code operates, let's go back and see how we could get the simpler building blocks, and then examine ideas about how they could have ever got together to form a system of very large polymers interacting in such a complex way. So we want to discuss the conditions on the primitive Earth, laboratory experiments (beginning with Miller-Urey) that produce interesting molecules that might be related to the beginning of biosynthesis, and whether the early Earth's conditions could have been "reducing" enough (containing H-rich compounds) for this kind of chemical synthesis.