Nucleic acids

A1.2 · SL

How can four bases store so much information?

Structure protects a variable sequence

A small alphabet can form many sequences. Complementary pairing allows a sequence to be copied and used.

This unit connects nucleotides, DNA/RNA structure, coding and evidence. Sequence examples and numerical data are invented.

Prerequisites: covalent bonds, hydrogen bonds and condensation reactions.

A base is only part of a nucleotide

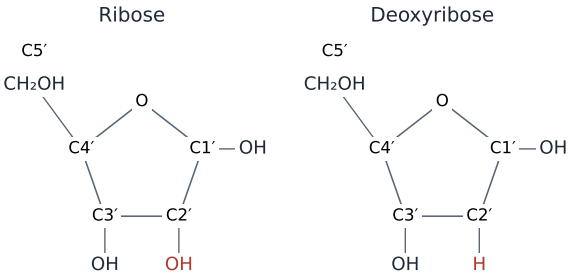
Keep three components connected correctly

A nucleotide has a phosphate group, a pentose sugar and a nitrogen-containing base. Draw a circle, pentagon and rectangle.

Connect phosphate to sugar and base to sugar. A base is not a whole nucleotide or part of the alternating backbone.

DNA is genetic material in living organisms. Some viruses have RNA genomes; RNA also has essential cellular functions.

Compare the sugars at the correct carbon



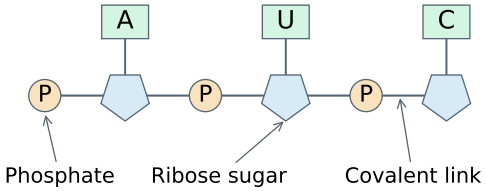
2′: OH in ribose; H in deoxyribose.
Other H and stereochemistry omitted.

Ribose has OH at 2′; deoxyribose has H there. Both sugars still contain oxygen.

Name the bases before reading the code

Nucleic acid	Bases
DNA	Adenine, thymine, cytosine, guanine
RNA	Adenine, uracil, cytosine, guanine

Trace one RNA chain



Continuous sugar–phosphate backbone.
Each base attaches to a sugar.

> Phosphodiester links form a continuous covalent sugar–phosphate backbone; bases attach to sugars.

Polymerization creates a strand

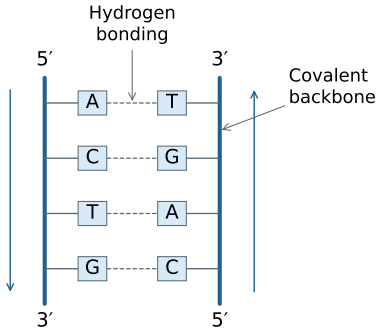
A condensation model has a stated boundary

Linking nucleotide units creates covalent sugar–phosphate connections. Three units need two new links in a simple linear model.

The backbone is not held together mainly by hydrogen bonds between successive bases.

Cells use activated nucleotide substrates. A classroom water-release diagram is not a complete intracellular reaction mechanism.

Distinguish along-strand and across-strand bonds



Backbones are covalent; hydrogen bonding pairs A with T and C with G between opposite strands.

A matching strand is complementary, not identical

Pairing constrains a new sequence

A–T–G–C–G–A pairs with T–A–C–G–C–T in the aligned opposite strand.

DNA normally forms a double helix with backbones outside and paired bases inside. A flat ladder can represent the pairing.

Antiparallel strands run in opposite directions; a diagram must show this relationship.

Count base pairs and nucleotides separately

1. **Invented molecule.** 800 base pairs; adenine is 28% of all nucleotides.
2. **Total and pairing.** $N = 2(800) = 1600$; A = T = 28%, C = G = 22%.
3. **Counts.** $N_A = N_T = 448$ and $N_C = N_G = 352$; sum = 1600.

Decision

A = T and C = G does not imply all four proportions are equal or describe each single strand.

Compare typical cellular nucleic acids

Feature	DNA versus RNA
Sugar	Deoxyribose versus ribose
Bases	A/T/C/G versus A/U/C/G
Usual strands	DNA: paired double strand; RNA: one often folded strand

RNA has several roles

Folding does not necessarily make two strands

mRNA carries coding information used in protein synthesis.

tRNA helps match amino acids to codons; rRNA is a structural and functional part of ribosomes.

Complementary regions can pair within one RNA chain. The chain can fold while remaining one continuous strand.

DNA and RNA compared

DNA



deoxyribose; A T C G

two paired strands

RNA



ribose; A **U** C G

one strand, can fold

Both are chains of nucleotides

Templates make a sequence reproducible

Complementary pairing supports copying

In replication, each DNA strand can guide a complementary new strand after separation.

In transcription, DNA A directs RNA U; DNA T directs RNA A; C and G pair.

Complementarity is important, but accurate copying also requires enzymes and checking. Mutations remain possible.

Keep coding DNA, template and mRNA distinct

1. **Coding DNA.** Invented ATG–GAA–TTC–TAA corresponds to mRNA AUG–GAA–UUC–UAA.
2. **Given meanings.** AUG = methionine; GAA = glutamate; UUC = phenylalanine; UAA = stop.
3. **Peptide.** Three amino acids, then termination. A stop codon adds no amino acid.

Decision

The template pairs with RNA; the coding strand matches the corresponding RNA except T replaces U.

When can a length calculation work?

1. **Model.** A complete translated mRNA coding segment has 90 bases and one terminal stop.
2. **Codons.** $N_{codons} = 90/3 = 30$.
3. **Amino acids.** $L = 30 - 1 = 29$. For 63 bases, $L = 63/3 - 1 = 20$.

Decision

Exclude untranslated regions and introns; specify initiation and termination. A whole gene is not automatically all coding sequence.

Degeneracy changes a mutation prediction

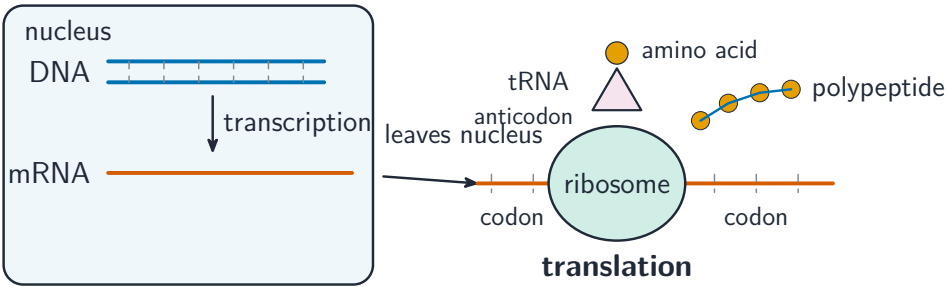
One amino acid can have several codons

GAA and GAG both encode glutamate. This illustrates degeneracy of the genetic code.

A GAA $\rightarrow$ GAG substitution can leave the amino-acid sequence unchanged.

Inserting one base within a translated region can shift subsequent codon grouping. The exact consequence depends on position and sequence.

Transcription and translation



tRNA brings the amino acids

A small alphabet gives many combinations

1. **Model.** Each of n specified positions has four possible bases: $N = 4^n$.
2. **Six positions.** $N = 4^6 = 4096$.
3. **Complement.** The second strand is constrained by pairing, not a second independent set of four choices.

Decision

Each finite molecule has finite capacity; varying sequence and length produces an enormous range.

A conserved code supports common ancestry

Same meanings do not require identical sequences

Codons generally have the same amino-acid meanings across living organisms.

This broad conservation supports inheritance of related information-processing systems from common ancestors.

Sequences can differ while the mapping remains shared. Known small code variants do not erase the general pattern.

How many possible sequences?



$n = 1$: $4^1 = 4$

$n = 2$: $4^2 = 16$

$n = 3$: $4^3 = 64$

$n = 6$: $4^6 = 4096$

$n = 20$: $4^{20} \approx 1.1 \times 10^{12}$

each position: A, T, C or G
$\Rightarrow$ multiply by 4 each time

the partner strand is fixed by pairing: it adds no extra choices

4^n grows very fast

Try a coding and base-count problem

Work before revealing the answer

An invented complete mRNA coding segment contains 72 bases including one terminal stop. Find peptide length under the stated model.

A double-stranded molecule has 500 base pairs and 24% adenine. Find each base count.

Then explain the bond types and assumptions behind the two calculations.

Reveal: count the correct objects

1. **Peptide.** $L = 72/3 - 1 = 23$ amino acids, assuming the complete coding region is counted.
2. **Total DNA.** $N = 2(500) = 1000$ nucleotides; A = T = 24%, C = G = 26%.
3. **Base counts.** $N_A = N_T = 240$; $N_C = N_G = 260$. Sum = 1000, giving 500 pairs.

Check yourself

Structure, sequence and evidence

Locate phosphate, sugar and base; distinguish covalent backbone links from hydrogen bonding across paired strands.

State whether a count concerns base pairs, nucleotides, one strand or the whole molecule.

Identify coding DNA, template and RNA before reading codons. A stop codon adds no amino acid; a whole gene is not automatically all coding.