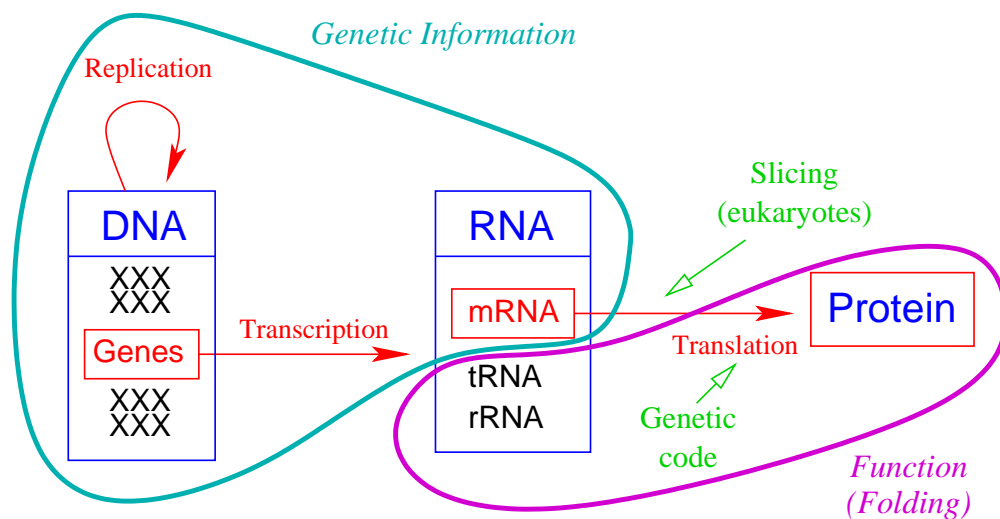


Molecular Biology's Central Dogma

DNA acts as a template to replicate itself, DNA is also transcribed into RNA, and RNA is translated into protein



- Linear polymer of nucleotides
- 4 nucleotides : Adenine, Cytosine, Guanine, Thymine
- Helical double stranded molecule
- Complementary and opposite strands

- Linear polymer of nucleotides
- Chemical backbone slightly different of DNA's one
- Single stranded molecule

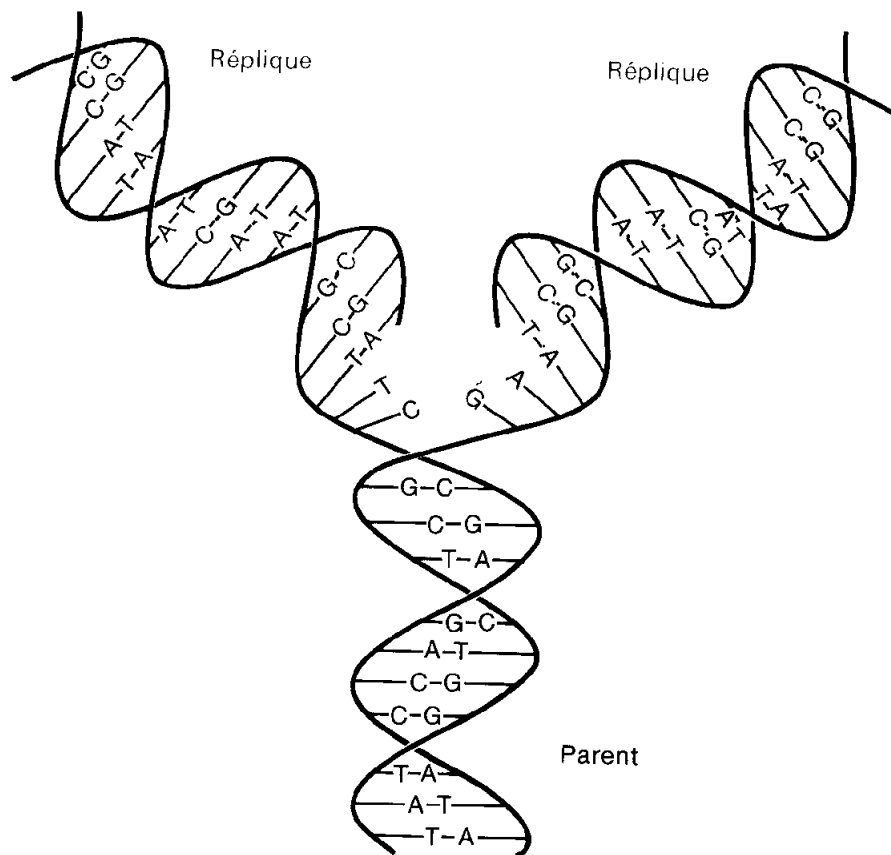
- Linear polymer of amino acids
- 20 different amino acids
- Each amino acid is determined by a trinucleotide (codon) of the mRNA (genetic code)

DNA Molecule

DesoxyriboNucleic Acid

- Macromolecule
- Linear polymer of *Nucleotides* A,C,G,T
- Genetic orientation : 5' → 3'
- Double helical structure
- **Complementary and opposite strands** : A-T , C-G
- Eukaryotes species : DNA in several *chromosomes* located in the nucleus of the cell
- Prokaryotes species : one DNA molecule
- Self-replication during cell division

DNA structure - DNA Replication



DNA sequence

A DNA sequence is a consecutive piece of one strand of a DNA molecule

- the DNA sequence is oriented : $5' \longrightarrow 3'$
- this strand determines completely the opposite strand
- all the information carried by this segment of the DNA molecule is contained in this DNA sequence
- this segment of the DNA molecule is completely defined by its sequel of letters A, G, C, T

\implies a DNA sequence \equiv a character string over the alphabet $\{A,C,G,T\}$

Example:

$s = \text{ATTTTCGATGGTGTCGTAGGTGTGCGATGTGCTGAT}$
can be a DNA sequence

DNA sequence analysis

The evolution affects DNA molecules (“mutations”) :

- point mutations : single nucleotide deletions, insertions, substitutions
- segment mutations : duplications, copies, deletions, ...

DNA sequence analysis tries to assert biological evidences from a deep analysis of the sequence of characters {A,C,G,T}:

1. Sequence comparison

- dot plot
- 2-2 sequence alignment
- multiple alignment
- databank similarity search
- classification
- etc...

2. Word composition analysis

3. Motif search

4. Location of repeats and approximate repeats

5. Nucleotidic composition analysis

6. Gene, introns, exons location

7. Informational content analysis

8. etc...