

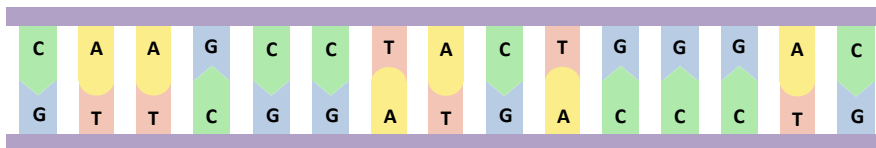
# 5.1

## DNA and protein synthesis

How DNA codes for proteins and the process of protein synthesis in ribosomes

### DNA, genes and polypeptides

DNA is a macromolecule formed when the single molecules called **nucleotides** polymerise into **nucleic acids**. There are two main types of nucleic acid: DNA and RNA. The main differences between the two are threefold: firstly, the pentose sugar in DNA is *deoxyribose*, whereas in RNA the sugar is simply *ribose*; in RNA the nitrogenous base uracil is present, replacing thymine of DNA; and lastly, RNA is a single helix whereas DNA is a double helix molecule, consisting of two intertwining and antiparallel helices.



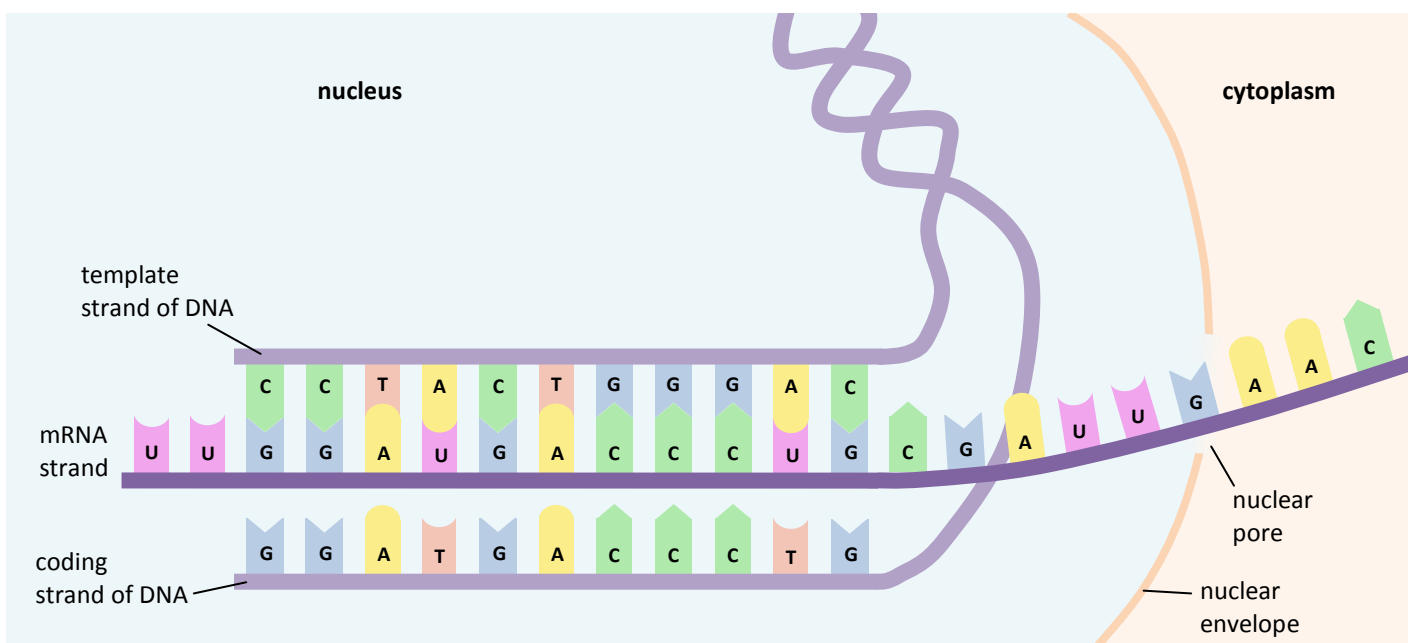
The diagram shows a length of DNA. The bases clearly interact in pairs: A matches to T, and C matches to G. All DNA has these base pairings.

A **gene** is a length of DNA which codes for a specific polypeptide, or in some cases, multiple polypeptides. This often means a gene codes for a protein, but not always: sometimes a protein is made of many polypeptides. The human **genome** comprises some three *gigabases* (three million base pairs), from genes within DNA. Most of the DNA is kept within the cell nucleus, but some DNA can be found in the mitochondrion (**mitochondrial DNA**).

The DNA sequence is a *triplet code*, whereby a short sequence of three bases codes for one **amino acid**. Although there are more than twenty possible arrangements of the four bases into groups of three, there are 20 amino acids which are coded for and used for **protein synthesis**. Most of the amino acids have multiple codes therefore that code for it. Some triplet codes do not code for an amino acid at all, but in fact indicate the end of the polypeptide chain, instructing the ribosome to stop the chain there (these are known as *stop codes*).

### Transcription

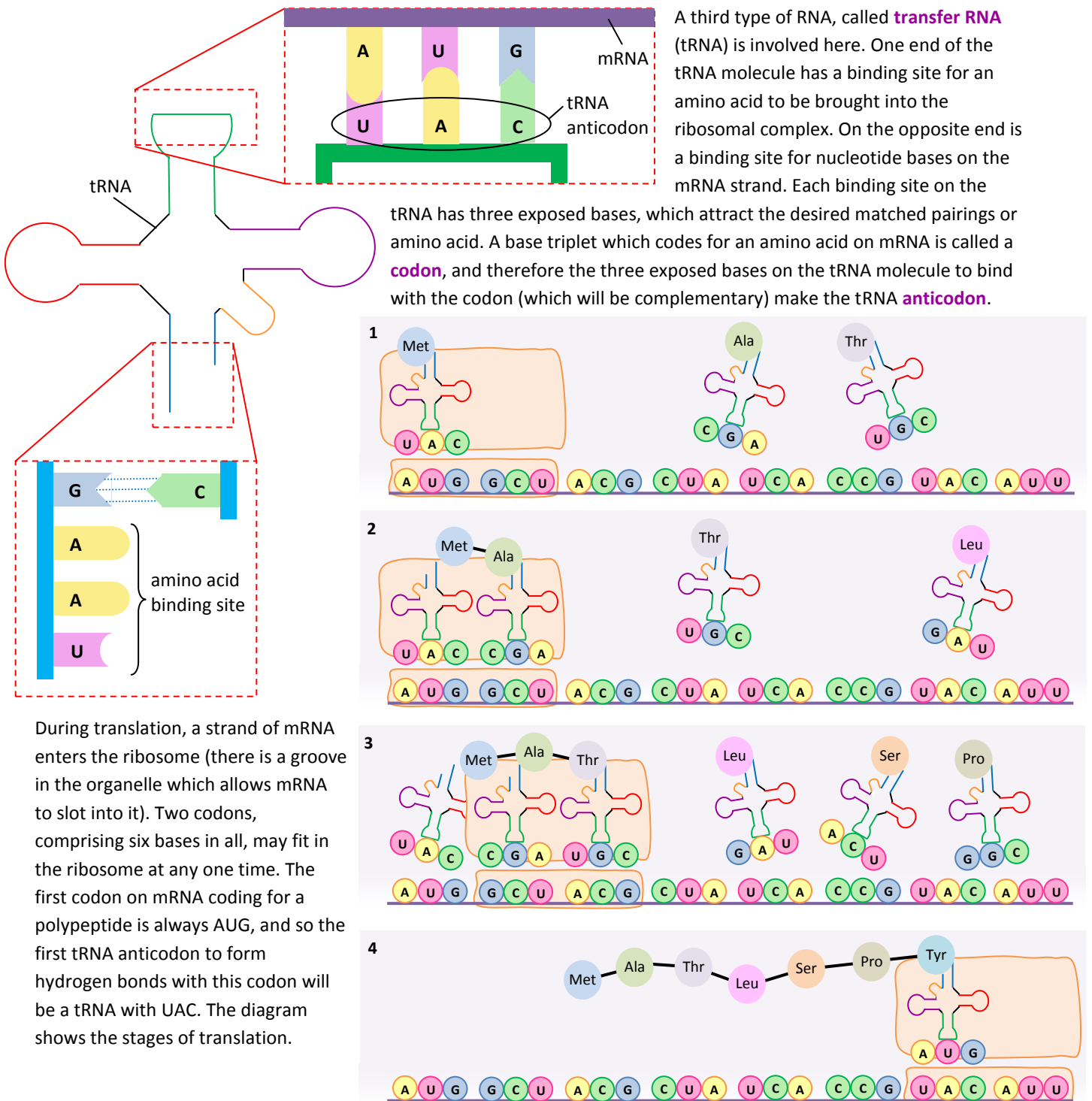
But how does the sequence of DNA from inside the nucleus result in production of a protein? The DNA double helix is far too large to leave the nucleus to code for protein synthesis, so a genetically-identical copy of the DNA has to be produced which can leave the nucleus. The nucleic acid **messenger RNA** (mRNA) is produced which is single-stranded and is therefore able to fit through **nuclear pores** and enter the cytoplasm to reach the ribosome. It is **transcription** which is the process that takes DNA and creates mRNA. The transcribed mRNA strand is identical to the **coding strand** of DNA in terms of base sequence, except **uracil** (U) replaces thymine as a base. The **template strand** provides the base sequence for mRNA, as the bases which line up will match the order of the coding strand.



- Before transcription can begin, the DNA double helix unzips at the gene to be expressed as a polypeptide
- The **hydrogen bonding** which holds together the complementary bases of DNA is broken to separate the helices
- There are two strands of DNA: the *coding strand* is the one which will be copied to be sent to the ribosome, and the *template strand* (which complements the coding strand) is used as a template for the mRNA
- Activated RNA nucleotides free in the nucleus bind to their exposed complementary bases of the template strand, catalysed by **RNA polymerase**, so the **cistron** (part of the gene to be converted into a polypeptide) is accurately copied
- The mRNA strand generated moves out of the nucleus via a nuclear pore and enters the cellular cytoplasm

## Translation

The second step in protein synthesis is **translation**. This is where the mRNA produced will ultimately code for a protein. Translation involves providing information for the correct sequence of amino acids to be brought together to create the specified polypeptide or polypeptides. The site of translation is in the **ribosomes** (either free in the cytoplasm, or bound to the rough **endoplasmic reticulum** surrounding the nucleus). Ribosomes are made of protein and ribosomal RNA, rRNA.



- 1** A ribosome begins as two separate subunits: known as the *30-s ribosome* (smaller subunit) and the *50-s ribosome* (larger subunit), so when protein synthesis is to occur the two join together to form what is known as *70-s ribosome* (commonly just known as a ribosome). The mRNA molecule attaches itself to the smaller subunit, leaving the codons exposed to the larger subunit, the first codon always being AUG (for *methionine*). The appropriate tRNA with the correct anticodon, UAC, arrives and enters the first of two spaces in the ribosome, bringing with it the amino acid
- 2** The second space in the ribosome allows a second tRNA to enter, and bind to the exposed codon on the mRNA (in the case of the diagram, the codon is GCU, so the tRNA brings the anticodon CGA and the amino acid *alanine*). A peptide bond forms between the two amino acids present, and the first dissociates from the tRNA
- 3** The ribosome moves along the mRNA strand one codon (three bases), and the first tRNA leaves the site, leaving its amino acid behind. The next tRNA arrives which correctly fits the complementary codon, and a peptide bond forms between the new and previous amino acids, now making a chain of three amino acids. The peptide bond formations in this process are catalysed by an enzyme present in the smaller *30-s* subunit
- 4** The ribosome continues to move along the mRNA strand, with a tRNA attaching to each exposed codon, and adding one amino acid to the polypeptide chain each turn. The chain grows until a stop codon is reached