

Protein-forming and non-protein-forming DNA sequences

DNA is protein-forming in humans only for a very small part of its length. As a gene we denote the section of DNA that codes for the formation of the final product. It can be a protein (formed from mRNA) or RNA, which is not transcribed into a protein (tRNA, rRNA etc). This article will continue to talk only about the formation of proteins.

A distinction needs to be made here between **protein-forming** and **encoding**. It has already been established that a large part of the „non-protein-forming genome“ is transcribed into RNA and, in addition to forming final tRNAs and rRNAs, can interact with molecules mRNA (siRNA, miRNA and interactions like them), therefore, probably also in a certain way intervene in the formation of proteins.

The gene contains in its transcribed part of the region directly encoding the sequence of amino acids of the protein (**exons**) even non-coding regions (**introns**) and untranscribed regions at the 3' end (polyadenylation signal) and at the 5' end (promotor). In addition gene expression they can affect distant sequences called **enhancers** and **silencers**.

The part of DNA related to protein formation (genes, including coding and non-coding parts, enhancers and silencers) occupies a total of about 20% of the length of the DNA in the human genome, the parts coding for proteins only about 3%. The remaining 80% is made up of DNA with an unclear function. Of this amount, approximately half is made up of repetitive sequences. ⚠

Repetitive sequence

🔍 For more information see *Repetitive sequences in the human genome*.

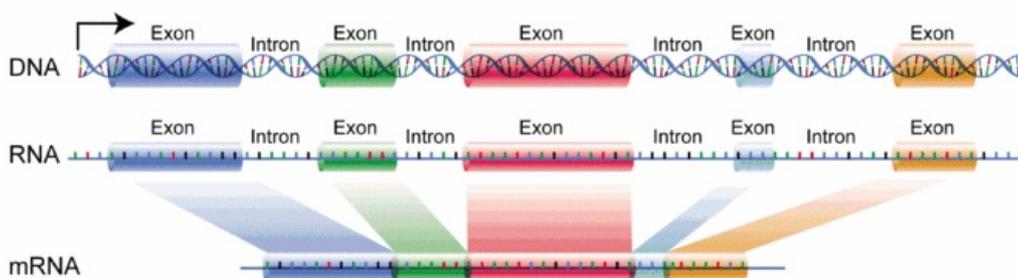
In terms of quantity in the genome, repetitive sequences are the most numerous, they make up approximately half of the DNA that is not related to the formation of proteins, i.e. approximately 40% of the entire length of the genome. These are parts of the genome in which certain sequences of nucleotides are regularly repeated.

Repetitive sequences can be divided into tandem and scattered sequences. **Tandem repetitions** are periodically repeating DNA sequences that are lined up close to each other. According to the length of the repeated sequence and the length of the total repeat, we distinguish between microsatellites, minisatellites and macrosatellites. **Scattered repetitions** are repetitive sequences that are scattered throughout the length of DNA due to re-copying and insertion at a different location. We divide them into *retrotranspososomes* (copying by reverse transcription) and *DNA transpososomes*. Retrotranspososomes include SINE and LINE sequences.

Protein-forming and non-protein-forming parts of a gene

The transcription of DNA into RNA is similar for all organisms, although the subsequent modifications already differ. In bacteria, the DNA is located directly in the cytoplasm, where the ribosomes are also located, and thus the translation of mRNA into protein occurs immediately. In eukaryotes, however, DNA is stored in the nucleus, from where it subsequently passes (after translation into RNA) through nuclear pores into the cytoplasm. Before the translation of this RNA into an amino acid sequence, post-transcriptional modifications occur - first, *capping* and *polyadenylation*. Next, splicing takes place - **introns" are cut out of the RNA (the parts that are not translated make up the majority of the gene). We refer to the remaining parts as exons.**

Exons - protein-forming sequences



DNA, exons and introns

Exons make up a relatively small part of DNA and nuclear RNA. It's probably just about 5% of the original sequence. After cutting out the introns, the remaining parts of the mRNA are joined together. This nuclear form reaches the ribosomes, where it is translated into individual proteins in the process of translation. This also implies the

importance of these sequences. When damage occurs in an intron region, there is usually no significant damage. However, any intervention in the exon region is absolutely essential and often leads to the formation of a defective or altered protein.

Introns - non-protein-forming sequences

Introns are part of both DNA and nuclear RNA, but they no longer reach the cytoplasm and ribosomes. Their information is not translated into proteins. It makes up the vast majority of the human genome (95%). Their size is different, often in between 80–10 000 nucleotides.

During the formation of mRNA, both introns and exons are transcribed. Before it leaves the nucleus, however, all introns are cut out and exons spliced together – the cytoplasmic mRNA is therefore significantly shorter. This modification is called **RNA splicing** – RNA splicing. The enzyme complex responsible for RNA splicing is called the spliceosome.

What is the recognition of introns for excision? Everything is decided by the nucleotides at both ends of the introns. Usually the first two nucleotides of an intron are GU and the last two are AG. Special enzymes take care of the removal – **snRNP** (small nuclear ribonucleoprotein particles). In addition to cutting out introns, they also ensure the joining of exons. Enzymes allow both ends of the intron to approach and the formation of a lasso-like structure, which is eventually removed.

Significance of introns

So what are the seemingly useless parts of DNA needed for? They played their role especially at the beginning of the evolution of genes, often speeding up the creation of new proteins using exon recombination. The second advantage is the possibility of the so-called "alternative editing" – instead of cleaving one intron, one exon may be "skipped", so that the cleaved region contains an intron, an exon, and the following intron, at the end of which the end of the splicing region is recognized. In this way, several different types of mRNA can arise based on the developmental stage of the cell. Thus, several different proteins can be produced from one gene. Everything depends on the connection of coding sequences – *exons*.

If mutations occur in introns, mostly in the first or last two nucleotides, the end or start of the intron may not be recognized by the spliceosome, so alternative splicing can occur. For example, in the BRCA protein, mutations of 18 nucleotides before the end of the intron, which cause aberrant splicing, are also known.

It is likely that there was a common ancestor of prokaryotes and eukaryotes that contained introns. However, in the course of evolution, the individual groups separated. Prokaryotes are characterized by a large number of divisions. A shorter genome (consisting only of exons) is therefore an advantage for them. Thus, the process of protein formation is accelerated. In eukaryotes, division does not occur so often, and therefore the non-coding parts of the genome have been preserved. On the other hand, a larger genome brings the advantage of possible recombination, a gene composed of introns and exons then has the possibility of alternative splicing.

3' a 5' untranslatable region of a gene

At both its ends, the gene contains regions that do not translate into a protein, but determine the level of its expression. At the 5' end of the gene there is a promoter, on which **transcription factors** and **DNA polymerase** sit. This region therefore determines the level of *gene expression*. At the 3' end of the gene there is a *polyadenylation signal* that affects the *termination of transcription*.

Links

Related articles

- Genome
- Transcription
- Translation
- mRNA
- Recombination
- Ribosome

External links

- RNA splicing (<https://www.ncbi.nlm.nih.gov/books/NBK28319/figure/A1709/?report=objectonly>)

References

- ALBERTS, B – BRAY, D – JOHNSON, A. *Basics of cell biology*. 2. edition. Espero Publishing, 2005. 740 pp. ISBN 80-902906-2-0.
- OTOVÁ, Berta. *Medical biology and genetics part I..* 1. edition. Karolinum, 2008. 123 pp. ISBN 978-80-246-1594-3.

