

# Eukaryotic gene structure

Eukaryotic genes consist of coding and noncoding segments of DNA, called exons and introns, respectively. At first glance it seems to be an unnecessary burden to carry DNA without obvious functions within a gene. However, it has been recognized that this has great evolutionary advantages. When parts of different genes are rearranged on new chromosomal sites during evolution, new genes may be constructed from parts of previously existing genes.

## Exons and introns

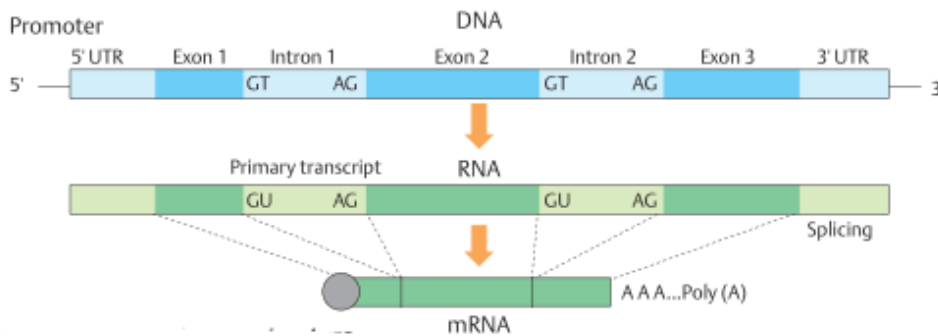
In 1977, it was unexpectedly found that the DNA of a eukaryotic gene is longer than its corresponding mRNA. The reason is that certain sections of the initially formed primary RNA transcript are removed before translation occurs. Electron micrographs show that DNA and its corresponding transcript (RNA) are of different lengths (1). When mRNA and its complementary single-stranded DNA are hybridized, loops of single-stranded DNA arise because mRNA hybridizes only with certain sections of the single stranded DNA. In (2), seven loops (A to G) and eight hybridizing sections are shown (1 to 7 and the leading section L). Of the total 7700 DNA base pairs of this gene (3), only 1825 hybridize with mRNA. A hybridizing segment is called an exon. An initially transcribed DNA section that is subsequently removed from the primary transcript is an intron. The size and arrangement of exons and introns are characteristic for every eukaryotic gene (exon/intron structure). (Electron micrograph from Watson et al., 1987).

## Intervening DNA sequences (introns)

In prokaryotes, DNA is colinear with mRNA and contains no introns (1). In eukaryotes, mature mRNA is complementary to

only certain sections of DNA because the latter contains introns (2). (Figure adapted from Stryer, 1995).

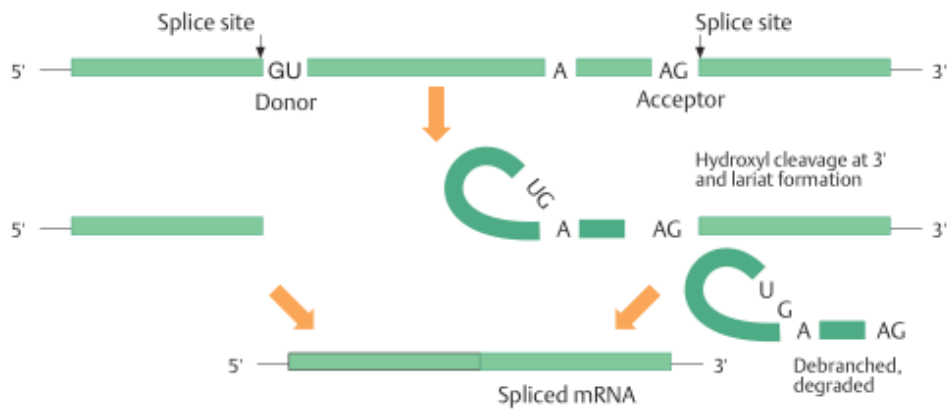
## Basic eukaryotic gene structure



### Basic eukaryotic gene structure

Exons and introns are numbered in the 5' to 3' direction of the coding strand. Both exons and introns are transcribed into a precursor RNA (primary transcript). The first and the last exons usually contain sequences that are not translated. These are called the 5' untranslated region (5' UTR) of exon 1 and the 3' UTR at the 3' end of the last exon. The non coding segments (introns) are removed from the primary transcript and the exons on either side are connected by a process called splicing. Splicing must be very precise to avoid an undesirable change of the correct reading frame. Introns almost always start with the nucleotides GT in the 5' to 3' strand (GU in RNA) and end with AG. The sequences at the 5' end of the intron beginning with GT are called splice donor site and at the 3' end, ending with AG, are called the splice acceptor site. Mature mRNA is modified at the 5' end by adding a stabilizing structure called a "cap" and by adding many adenines at the 3' end (polyadenylation).

## Splicing pathway in GU–AG introns



### Splicing pathway in GU – AG introns

RNA splicing is a complex process mediated by a large RNA-containing protein called a spliceosome. This consists of five types of small nuclear RNA molecules (snRNA) and more than 50 proteins (small nuclear riboprotein particles). The basic mechanism of splicing schematically involves autocatalytic cleavage at the 5' end of the intron resulting in lariat formation. This is an intermediate circular structure formed by connecting the 5' terminus (UG) to a base (A) within the intron. This site is called the branch site. In the next stage, cleavage at the 3' site releases the intron in lariat form. At the same time the right exon is ligated (spliced) to the left exon. The lariat is debranched to yield a linear intron and this is rapidly degraded. The branch site identifies the 3' end for precise cleavage at the splice acceptor site. It lies 18–40 nucleotides upstream (in 5' direction) of the 3' splice site. (Figure adapted from Strachan and Read, 1999)