- Exons: sequences found in mature RNA
- Introns: sequences removed from primary transcript (pre-mRNA)

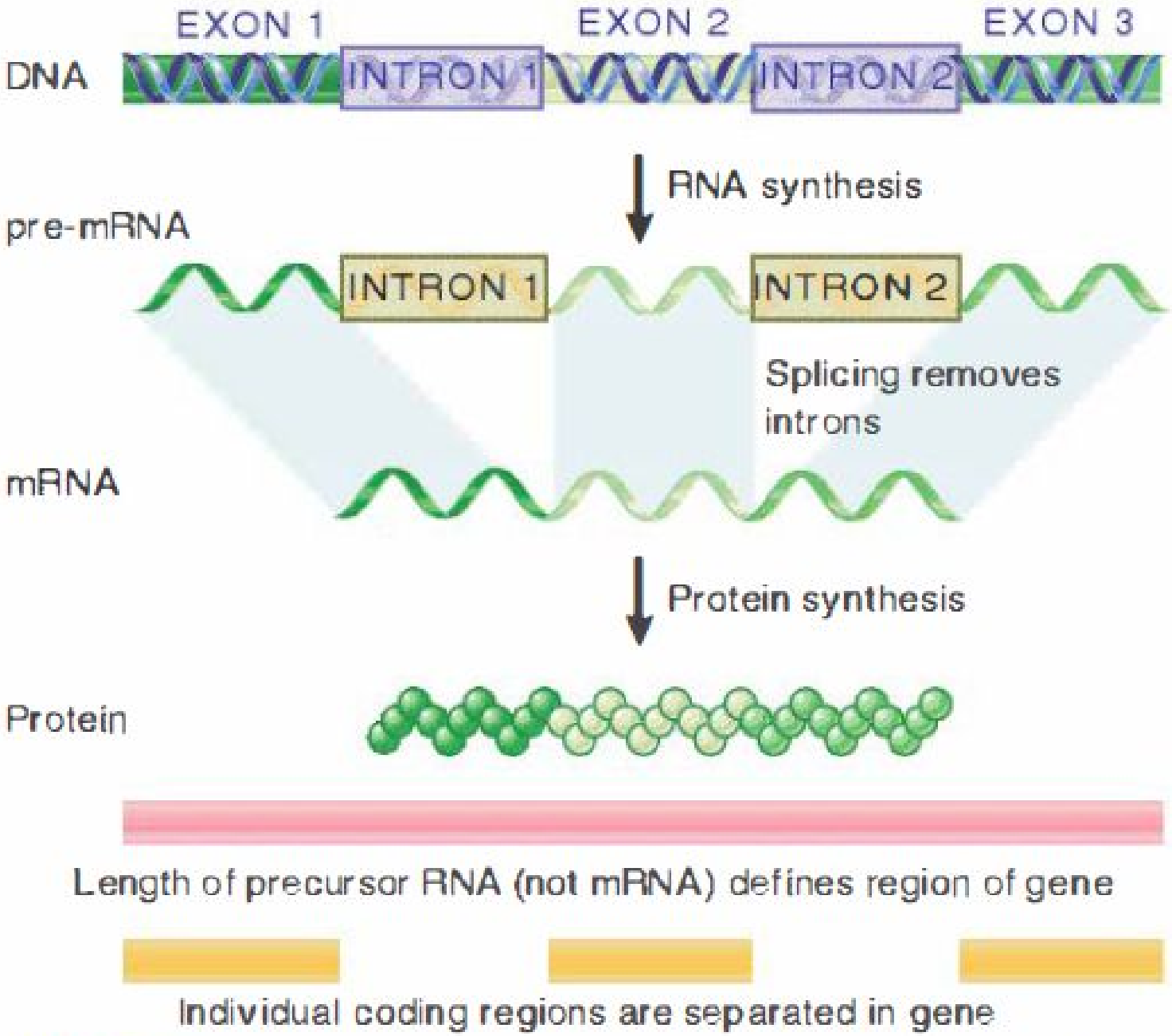


FIGURE 4.1 Interrupted genes are expressed via a precursor RNA. Introns are removed when the exons are spliced together. The mRNA has only the sequences of the exons.

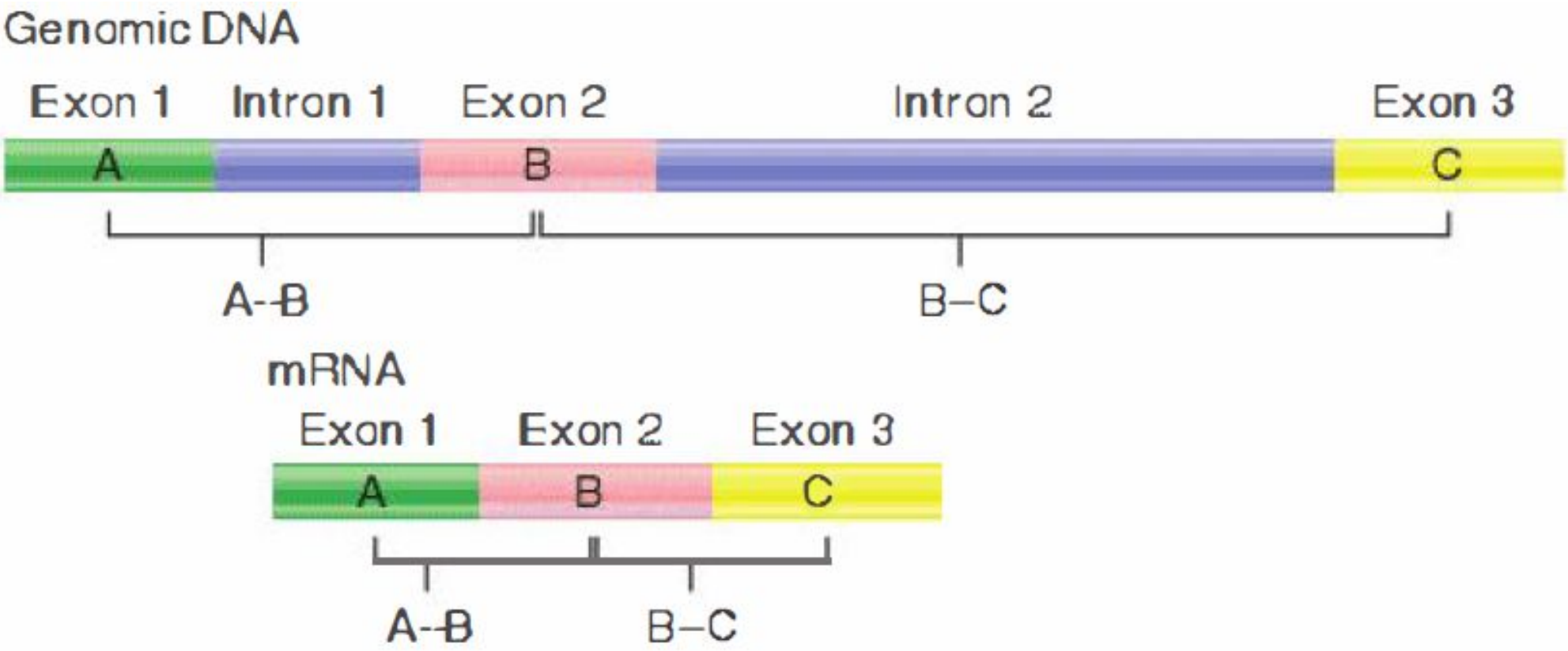


FIGURE 4.2 Exons remain in the same order in mRNA as in DNA, but distances along the gene do not correspond to distances along the mRNA or polypeptide products. The distance from A–B in the gene is smaller than the distance from B–C, but the distance from A–B in the mRNA (and polypeptide) is greater than the distance from B–C.

cDNA

Complementary DNA (cDNA) is DNA synthesized from a mature mRNA template in a reaction catalyzed by the enzyme Reverse Transcriptase.

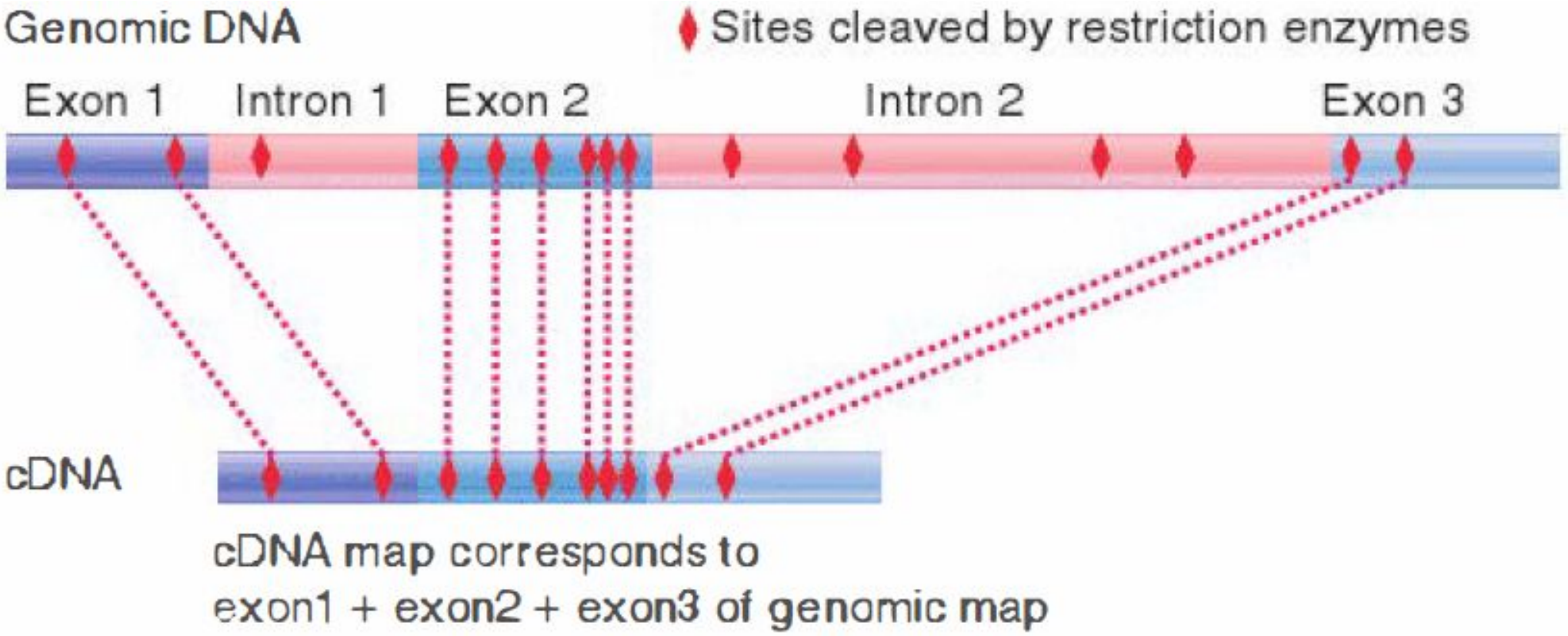


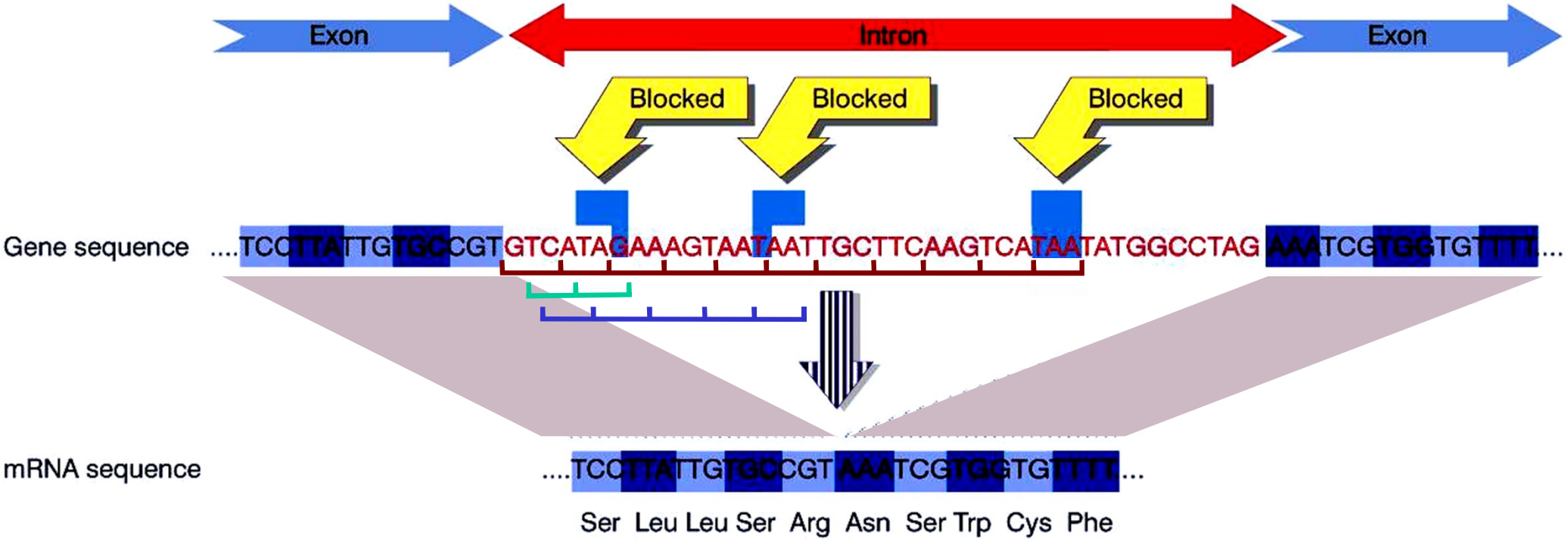
FIGURE 4.3 Comparison of the restriction maps of cDNA and genomic DNA for mouse β-globin shows that the gene has two introns that are not present in the cDNA. The exons can be aligned exactly between cDNA and the gene.

- Exons remain in the same order in mRNA as in DNA

- Nuclear genes introns usually do not encode proteins and all three possible reading frames are blocked by termination codons.

This means that if an intron is not correctly removed the translation of that mRNA is stopped.

Figure 2.12 An intron is a sequence present in the gene but absent from the mRNA (here shown in terms of the cDNA sequence). The reading frame is indicated by the alternating open and shaded blocks; note that all three possible reading frames are blocked by termination codons in the intron.



- Introns can be present in all gene types (pre-mRNA, tRNA, rRNA)

- In mRNA terminal exons include non translated regions (5' e 3' UTR)

Genes coding for the same protein in different species usually show the same exon-intron organization.

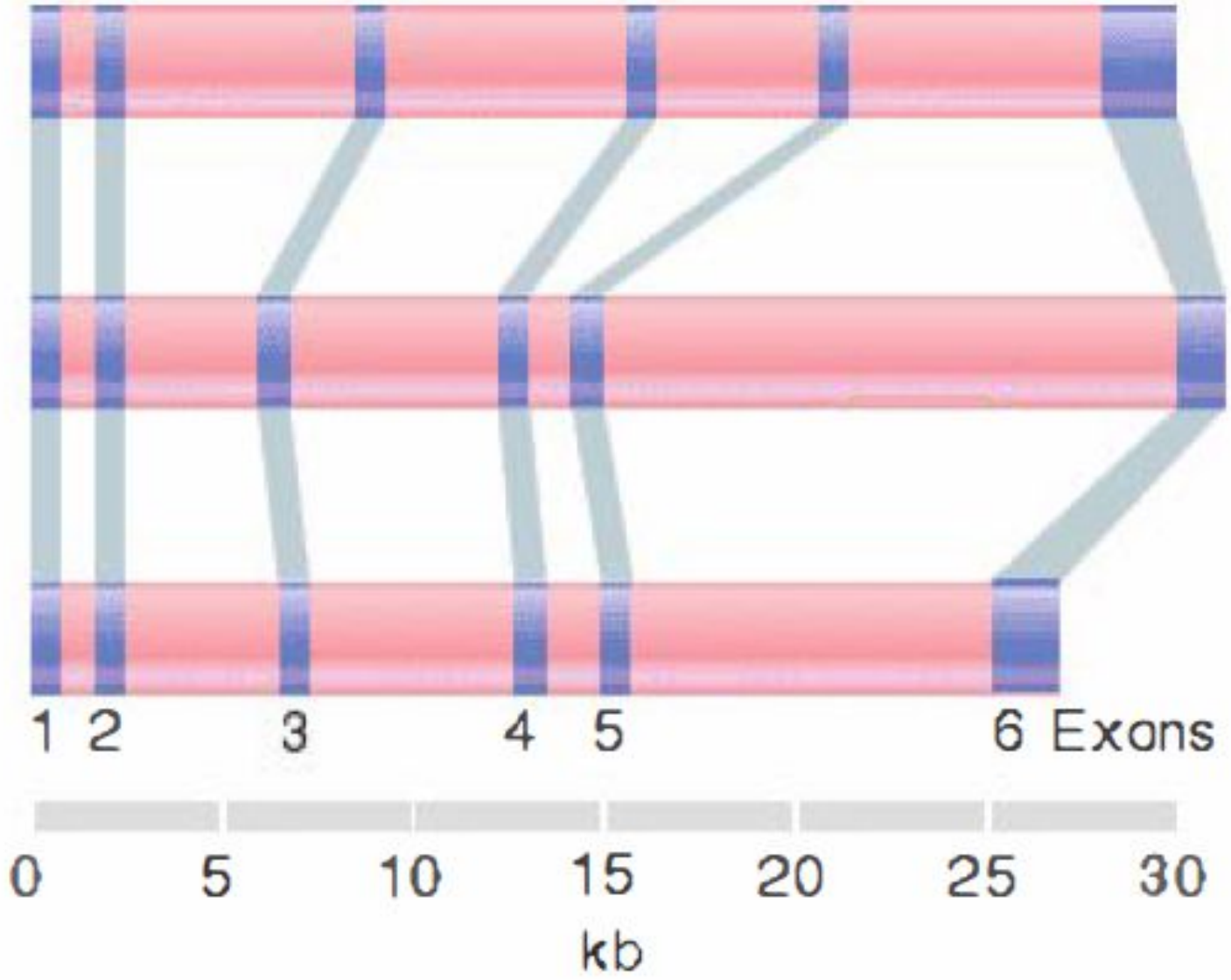


FIGURE 4.5 Mammalian genes for DHFR have the same relative organization of rather short exons and very long introns, but vary extensively in the lengths of introns.

DHFR → Dehydrofolate Reductase

Gene family → A gene family is defined as a group of genes that encode related or identical products as a result of gene-duplication events, starting from an ancestral common precursor gene.

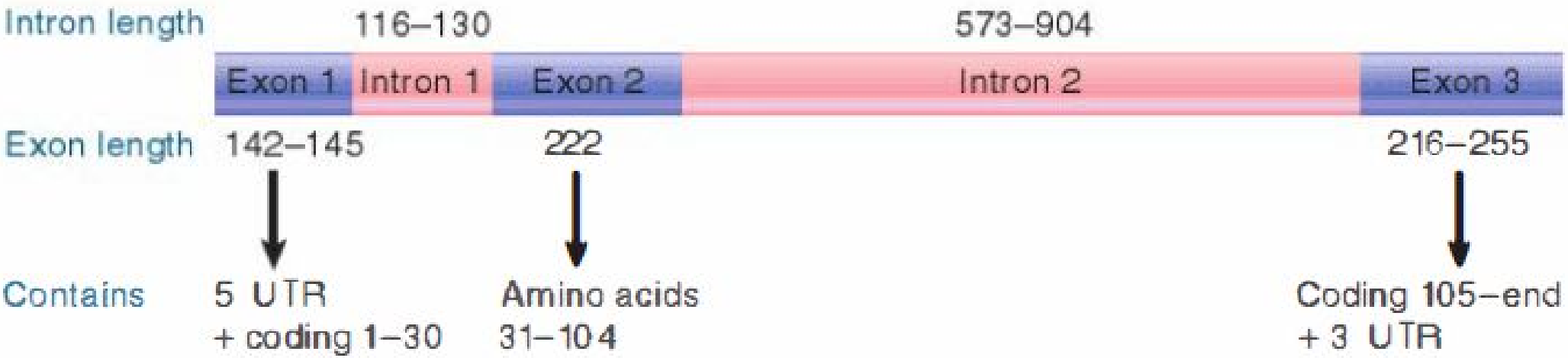


FIGURE 4.4 All functional globin genes have an interrupted structure with three exons. The lengths indicated in the figure apply to the mammalian β-globin genes.

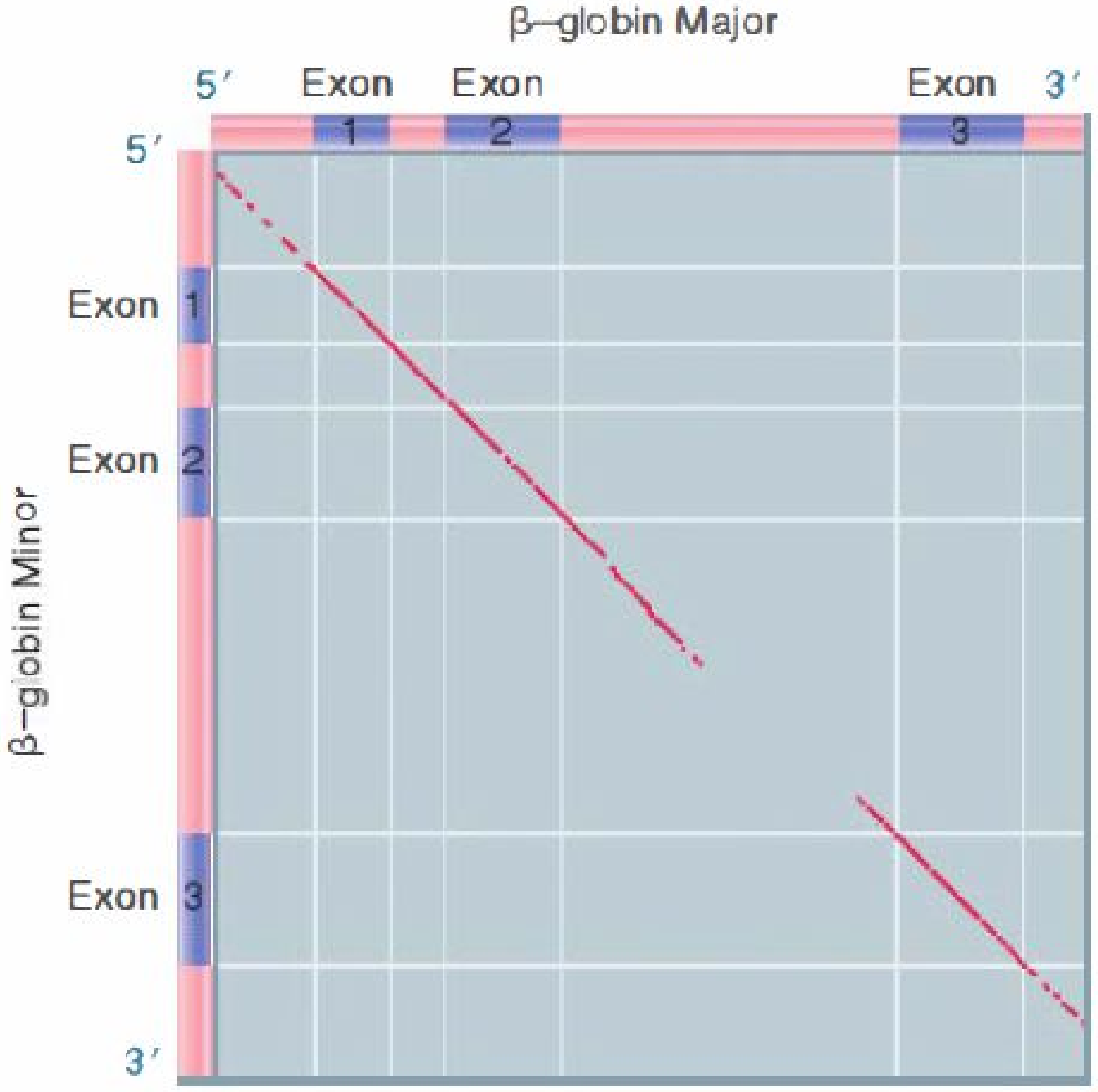


FIGURE 4.6 The sequences of the mouse β^{maj}- and β^{min}-globin genes are closely related in coding regions but differ in the flanking UTRs and the long intron. Data provided by Philip Leder, Harvard Medical School.

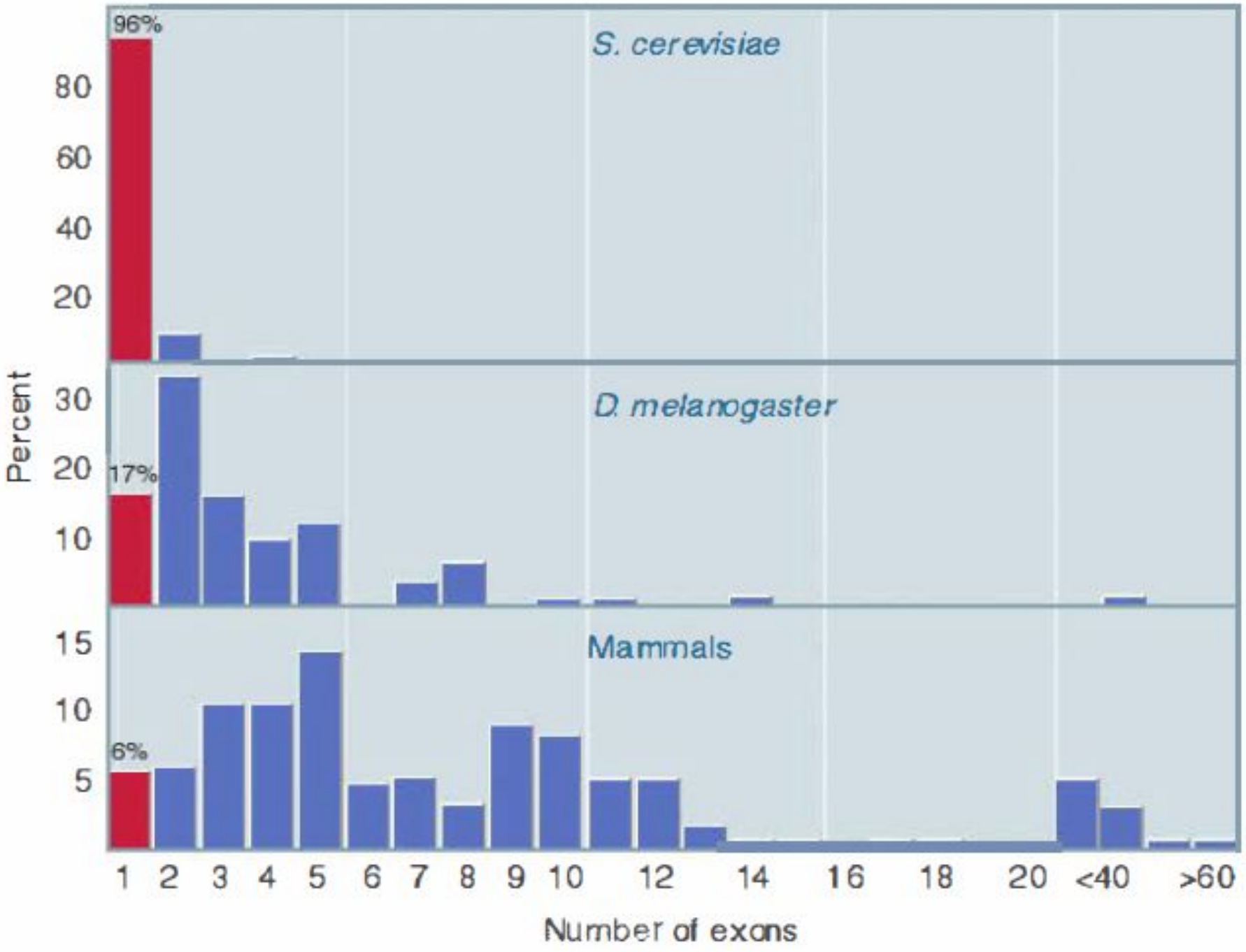


FIGURE 4.8 Most genes are uninterrupted in yeast, but most genes are interrupted in flies and mammals. (Uninterrupted genes have only one exon and are totaled in the leftmost column in red.)

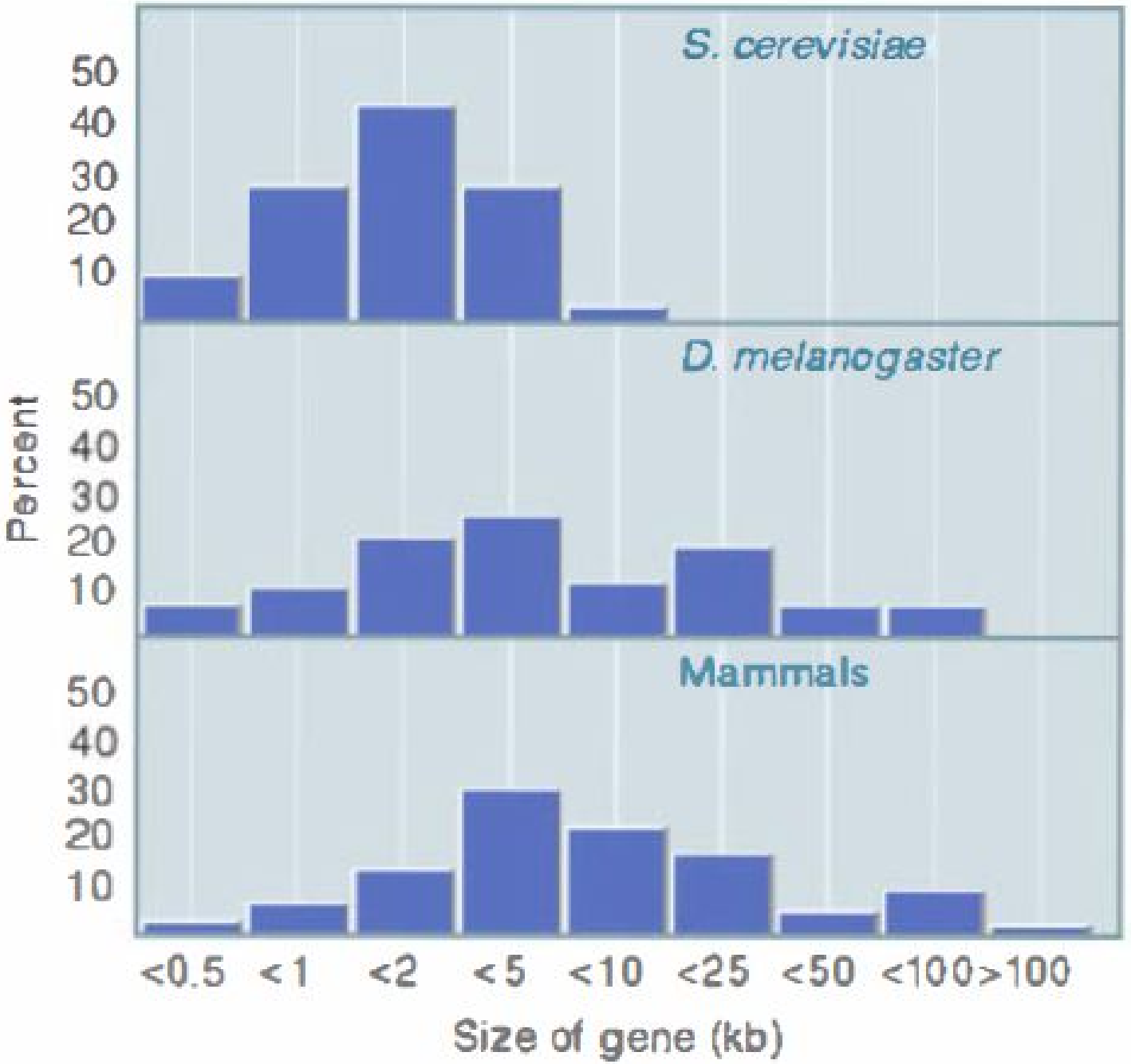


FIGURE 4.9 Yeast genes are short, but genes in flies and mammals have a dispersed bimodal distribution extending to very long sizes.

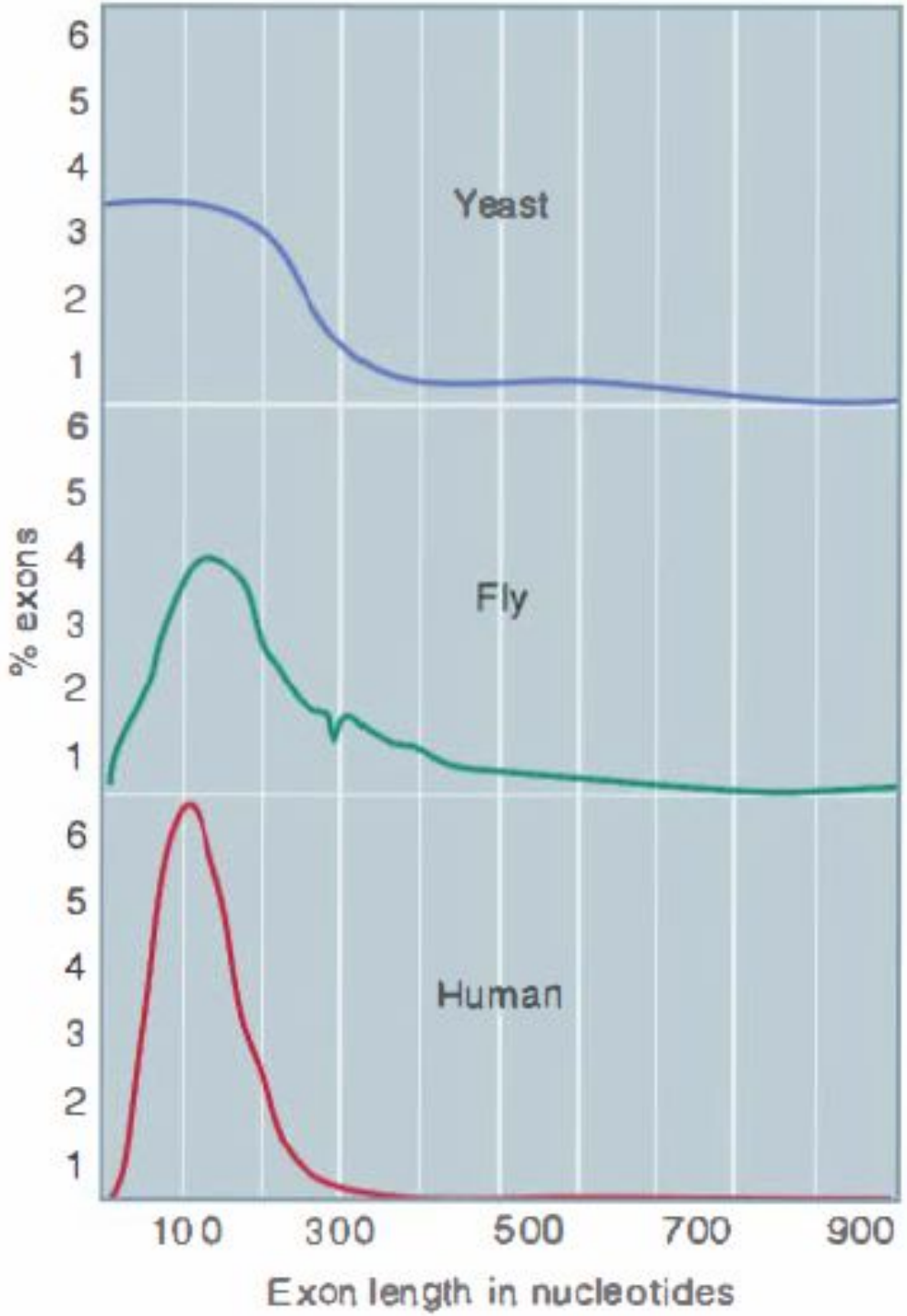


FIGURE 4.10 Exons encoding polypeptides are usually short.

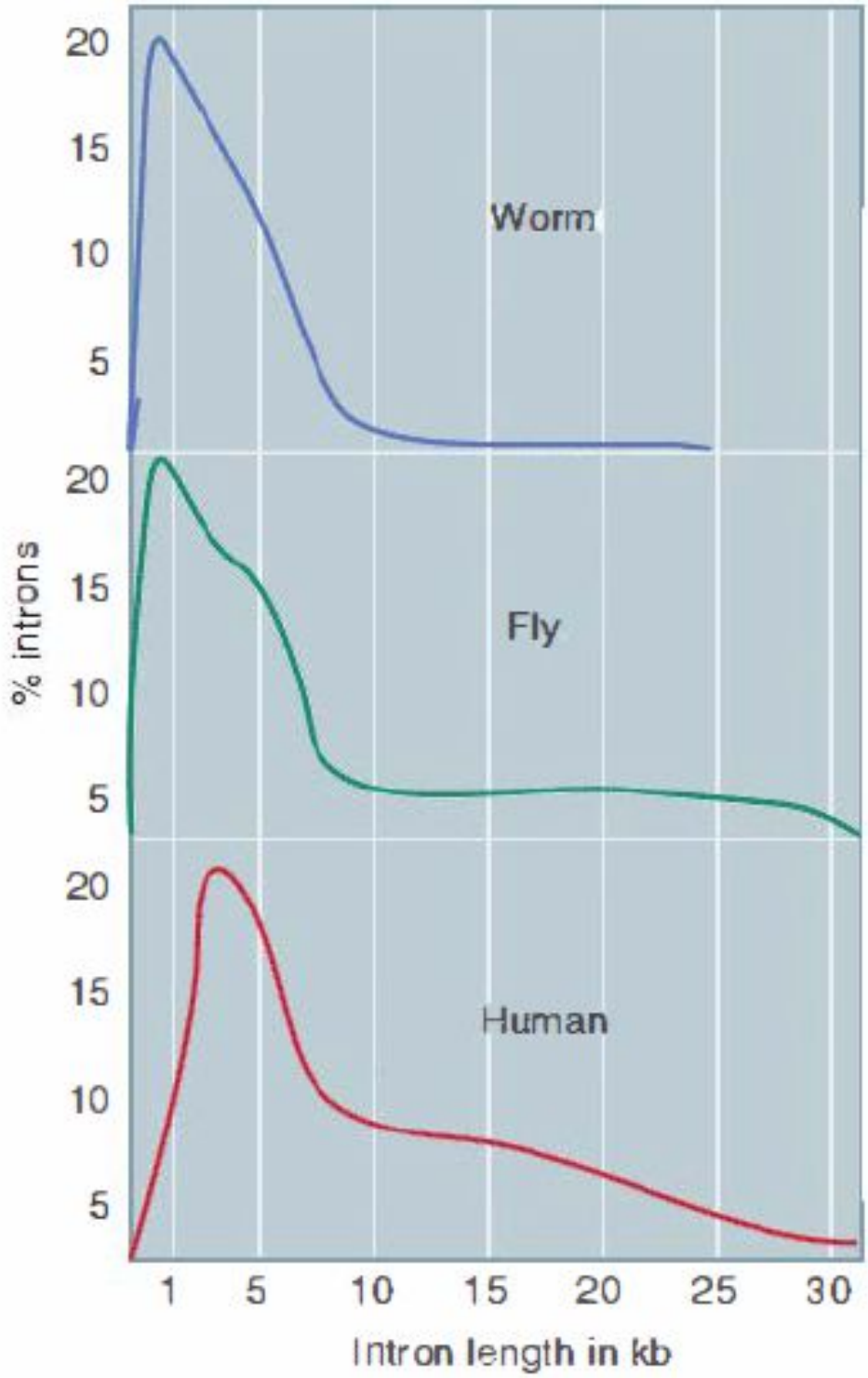


FIGURE 4.11 Introns range from very short to very long.

- Yeast: length of genes similar to length of messenger RNA.
- Mammals: length of genes about 5 times the length of mRNA.

Introns

- Absent in prokaryotes (few exceptions)
- Few in archaea and yeasts
- Many in complex eukaryotes

Origin of interrupted genes

"Introns early" hypothesis:
genes originated as interrupted structures and those now without introns have lost them in the course of evolution.

"Introns early" hypothesis:
the ancestral protein-coding sequences were uninterrupted and that introns were subsequently inserted into them.

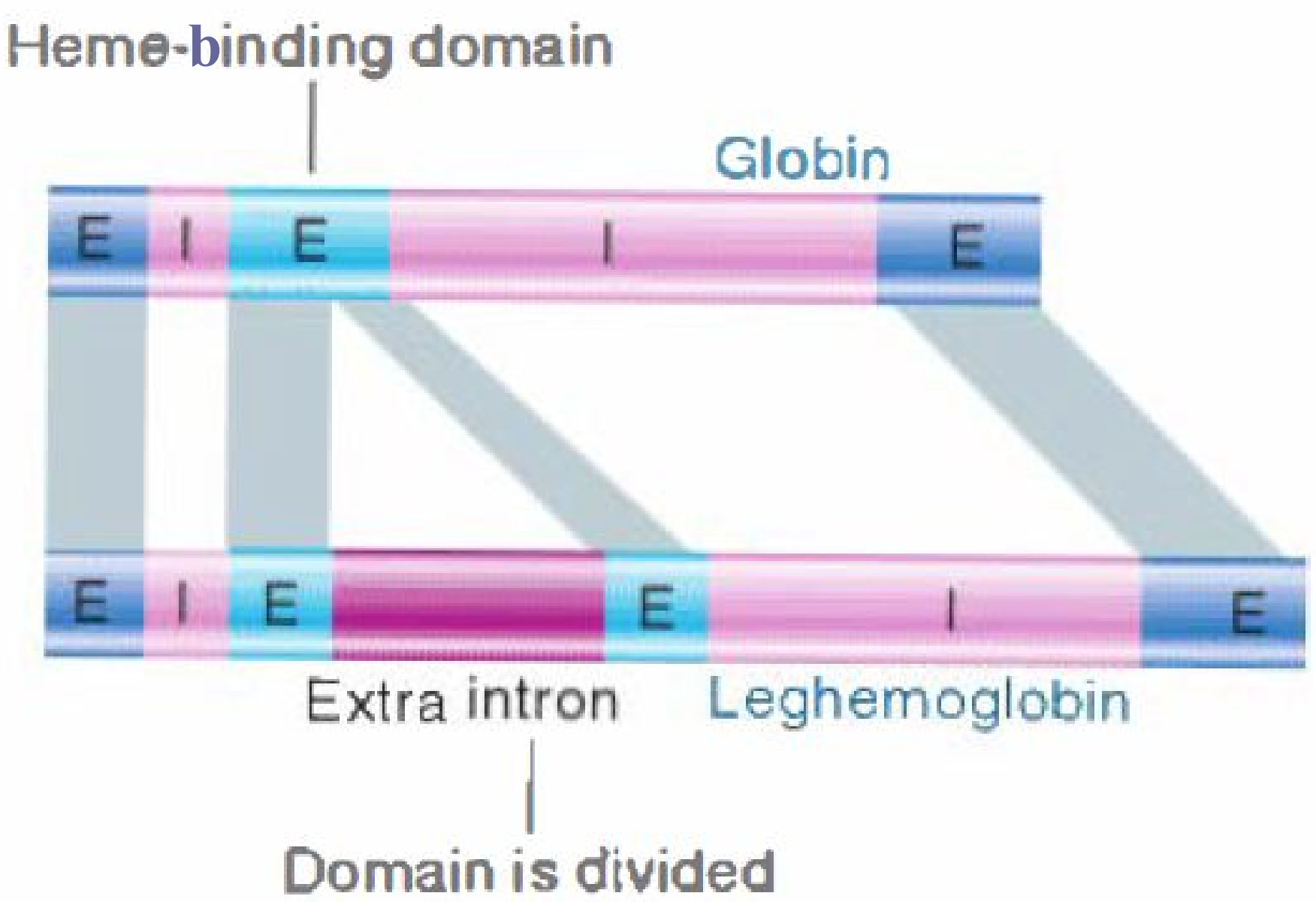


FIGURE 4.18 The exon structure of globin genes corresponds to protein function, but leghemoglobin has an extra intron in the central domain.

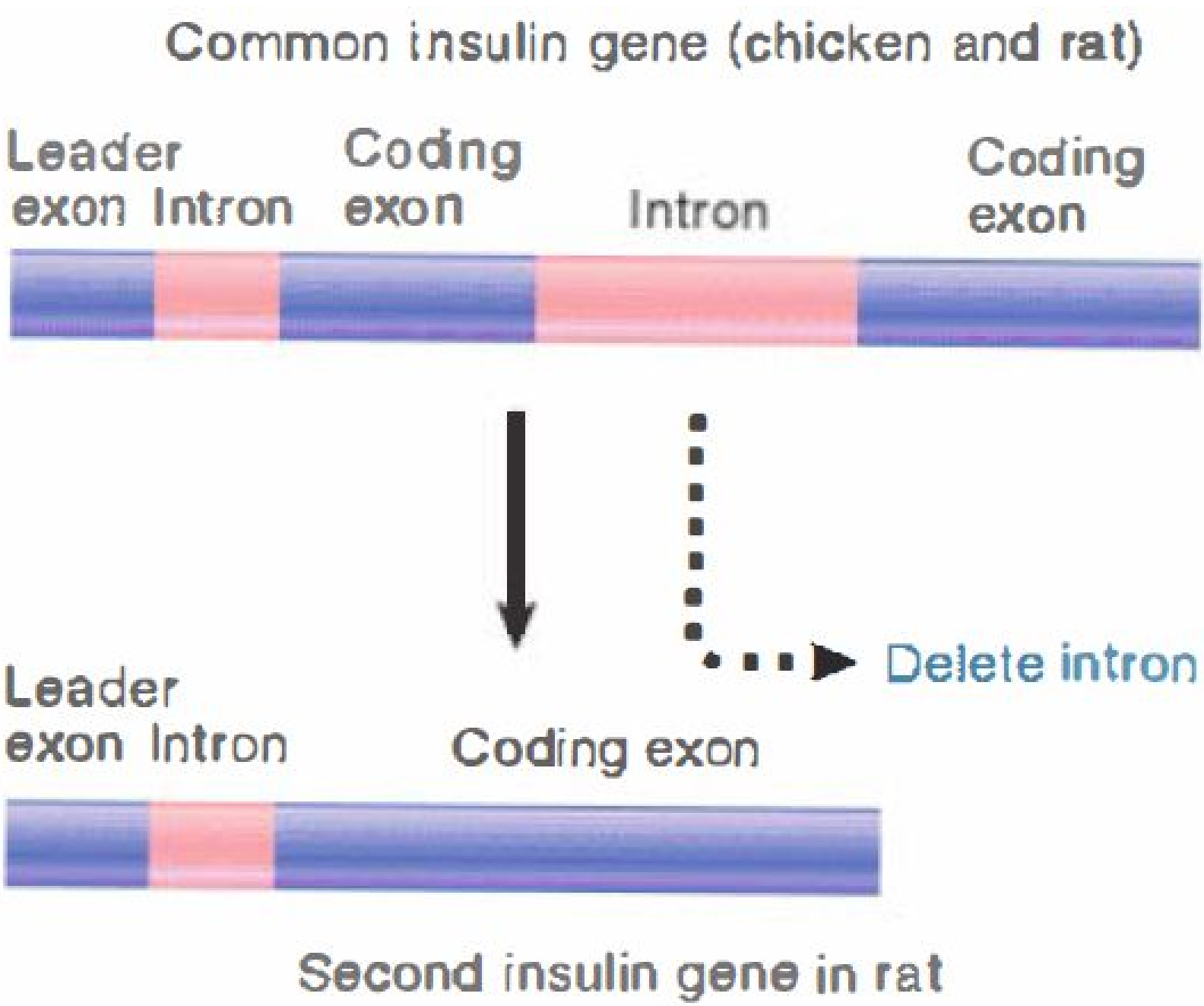
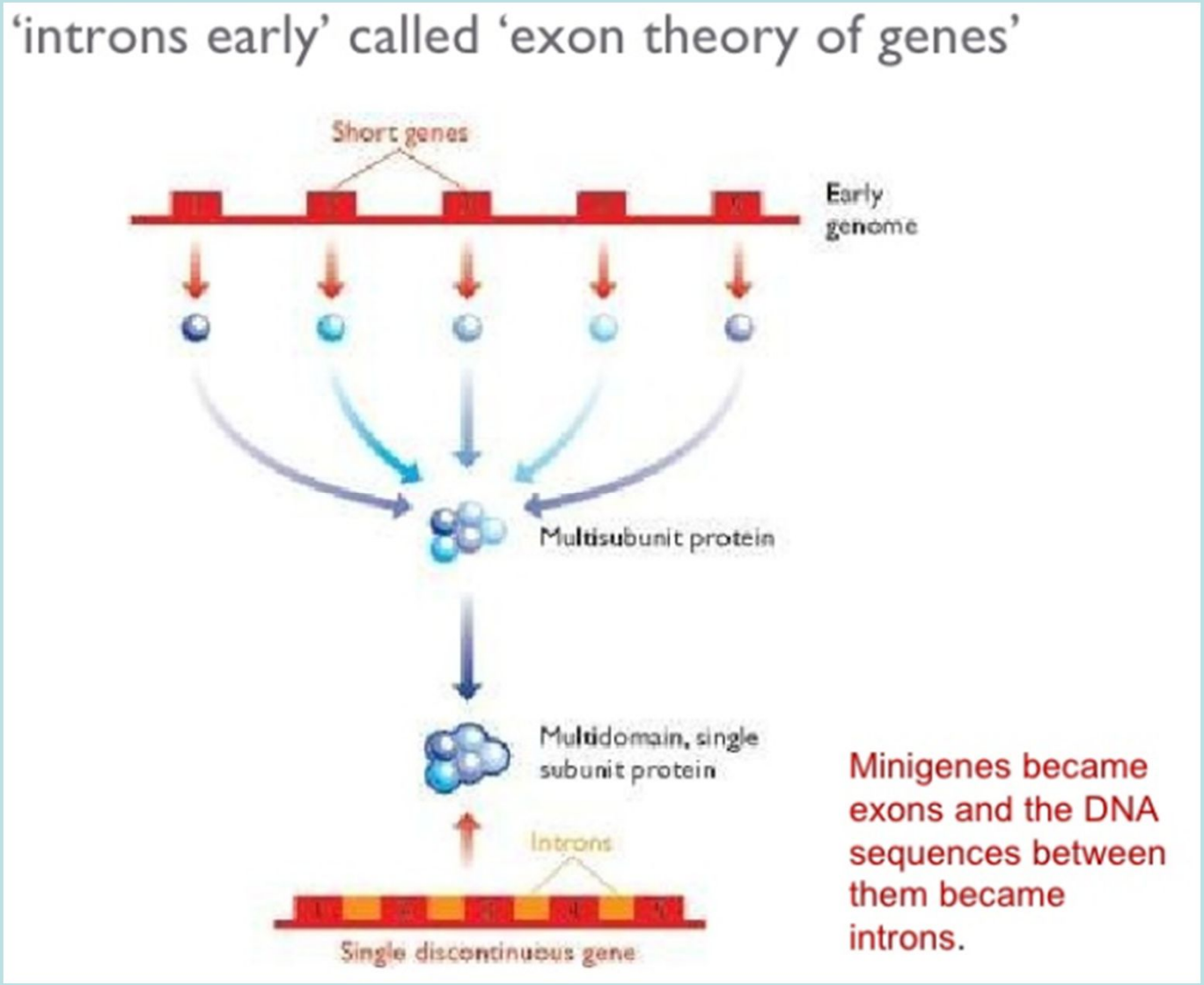


FIGURE 4.19 The rat insulin gene with one intron evolved by loss of an intron from an ancestor with two introns.

Exon theory of genes

The short genes forming ancients genomes probably encoded for single-domain proteins, and to have an active enzyme, they had to join together to form multi-subunits active complexes.

Later on during evolution this enzyme synthesis may have been made more efficient joining short genes, to form an interrupted gene coding for a single multi –domain protein subunit.



The evolutionary role of introns

(the "exon shuffling" theory)

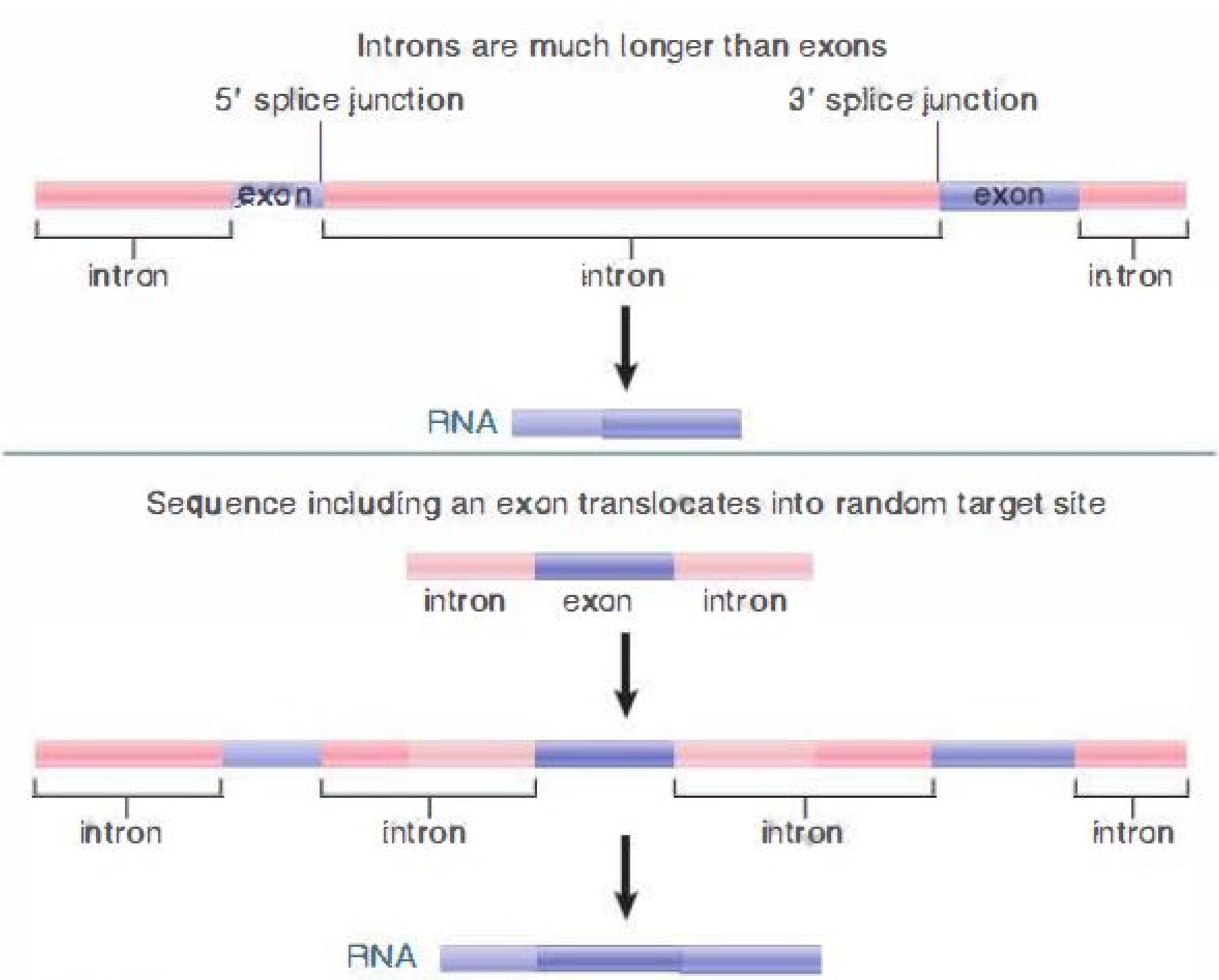
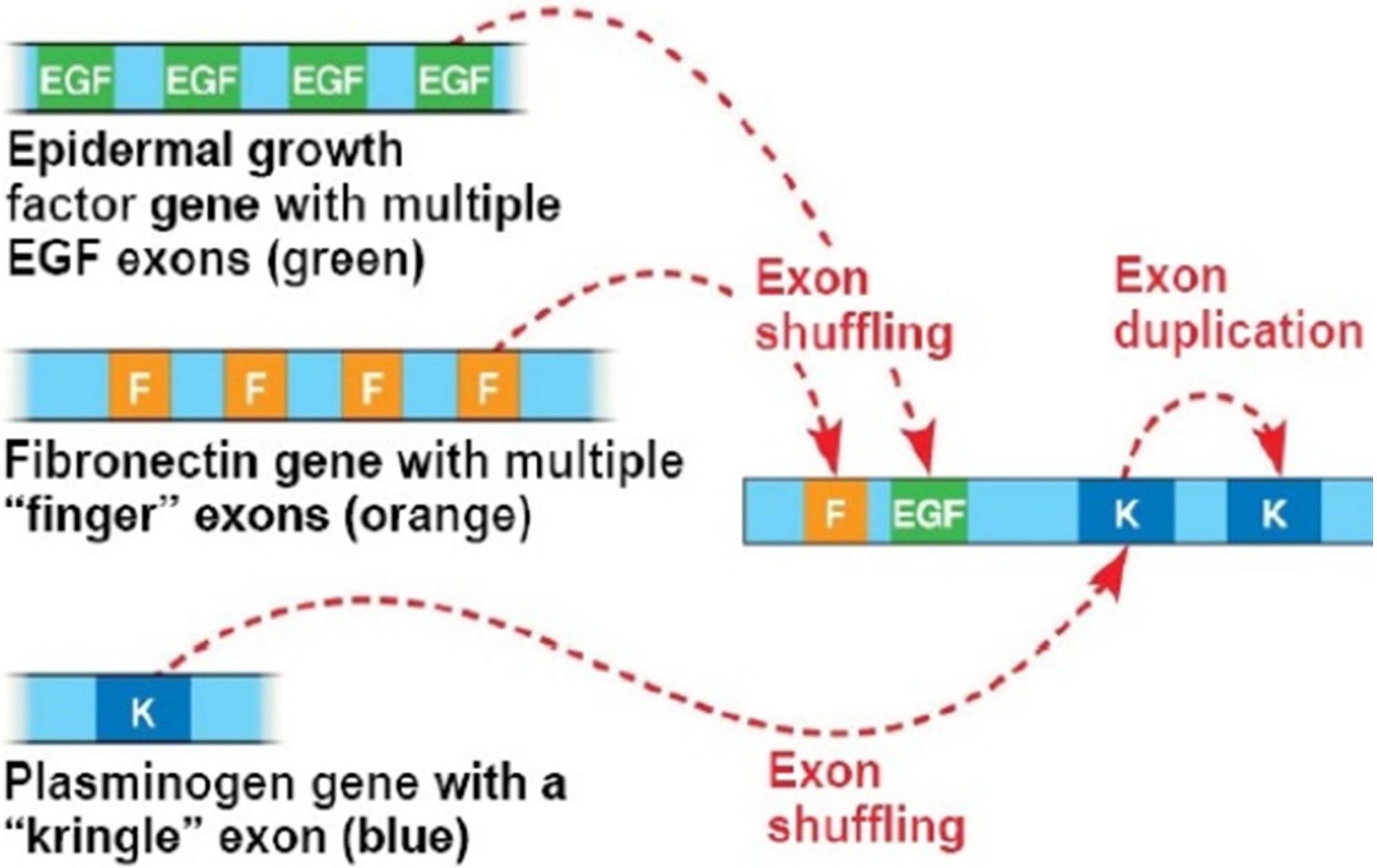


FIGURE 8.12 An exon surrounded by flanking sequences that is translocated into an intron may be spliced into the RNA product.

Domain shuffling: Tissue plasminogen activator (TPA),



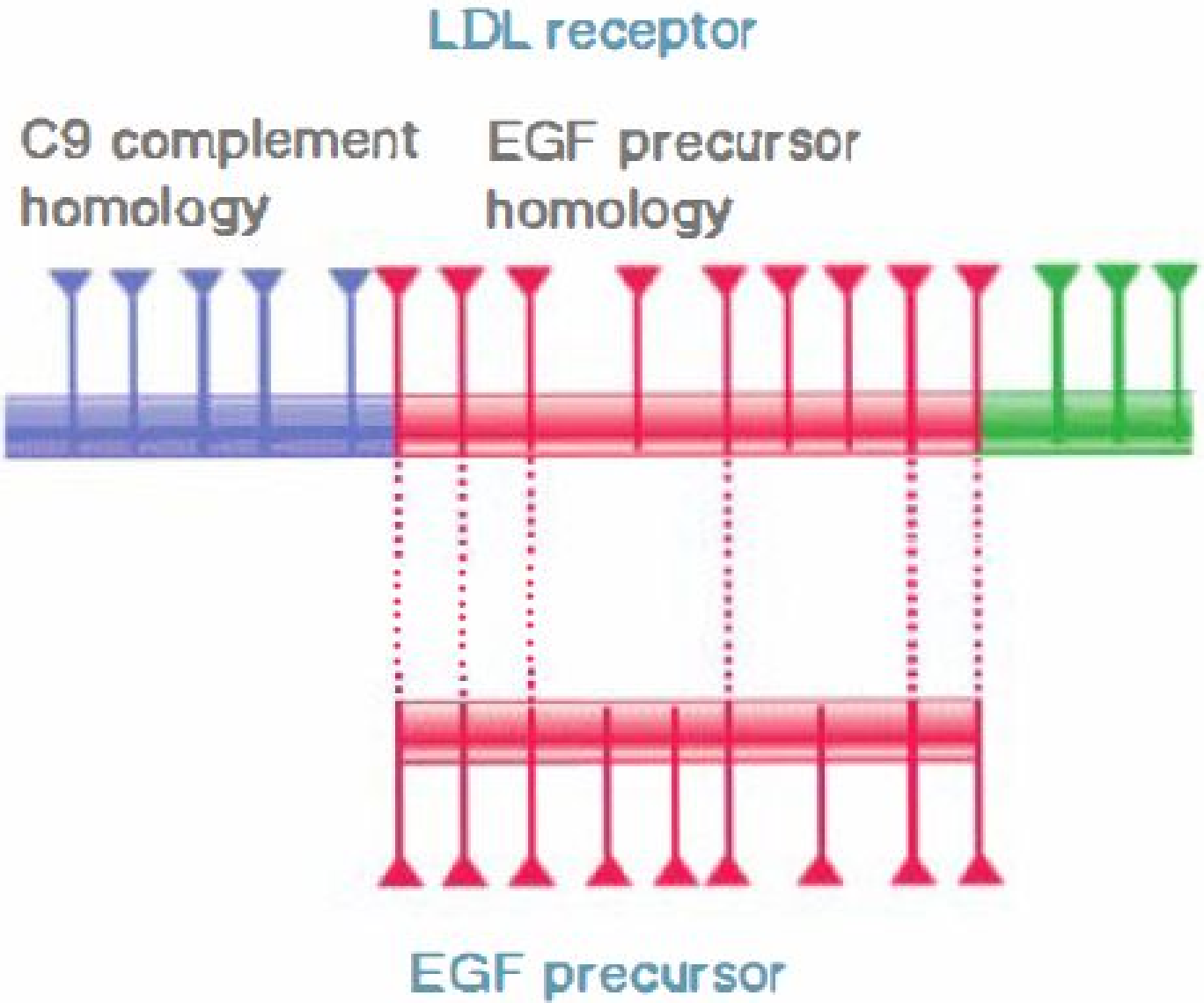


FIGURE 4.17 The LDL receptor gene consists of 18 exons, some of which are related to EGF precursor exons and some of which are related to the C9 blood complement gene. Triangles mark the positions of introns.

Modular structure of proteins

EGFP
Epidermal Growth
Factor Precursor

LDLR
Low Density
Lipoprotein Receptor

FN
Fibronectin

C9
Complement
Component 9

TPA
Plasminogen
Activator

UK
Urokinase

FX
Blood X
Coagulation Factor

Coding ability of eukaryotic genes

In eukaryotes (especially for higher eukaryotes) different proteins can be generated starting from a single gene containing partially overlapping sequences.

One gene ~~=~~ one protein

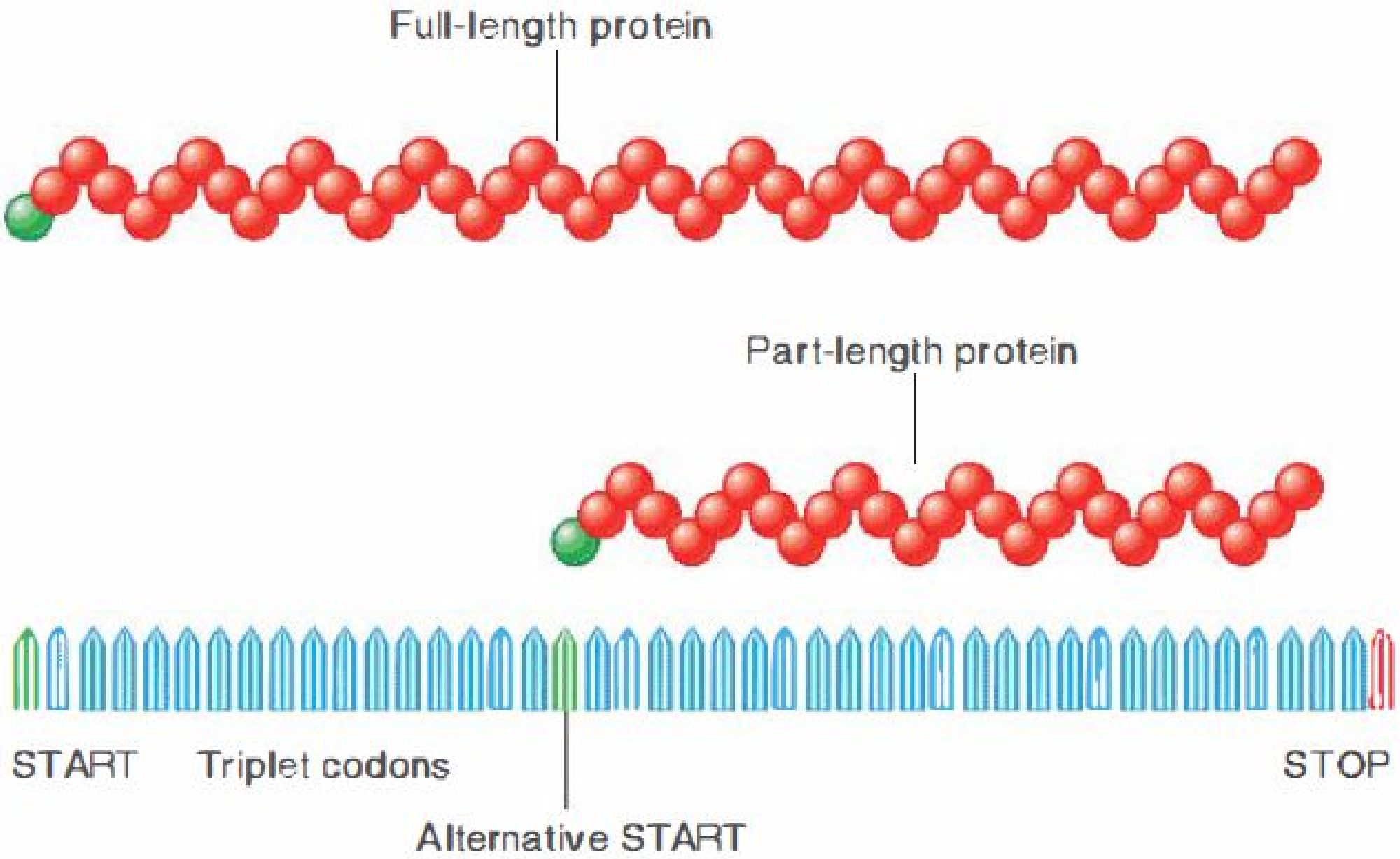


FIGURE 4.12 Two proteins can be generated from a single gene by starting (or terminating) expression at different points.

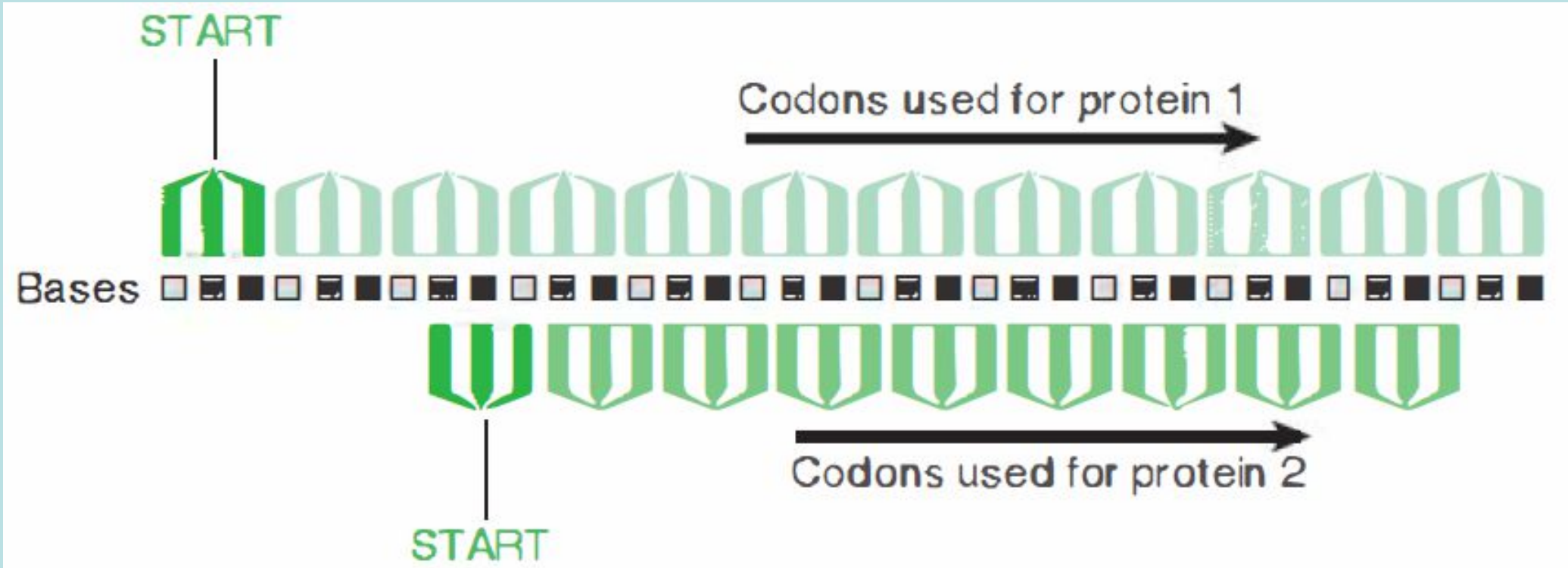
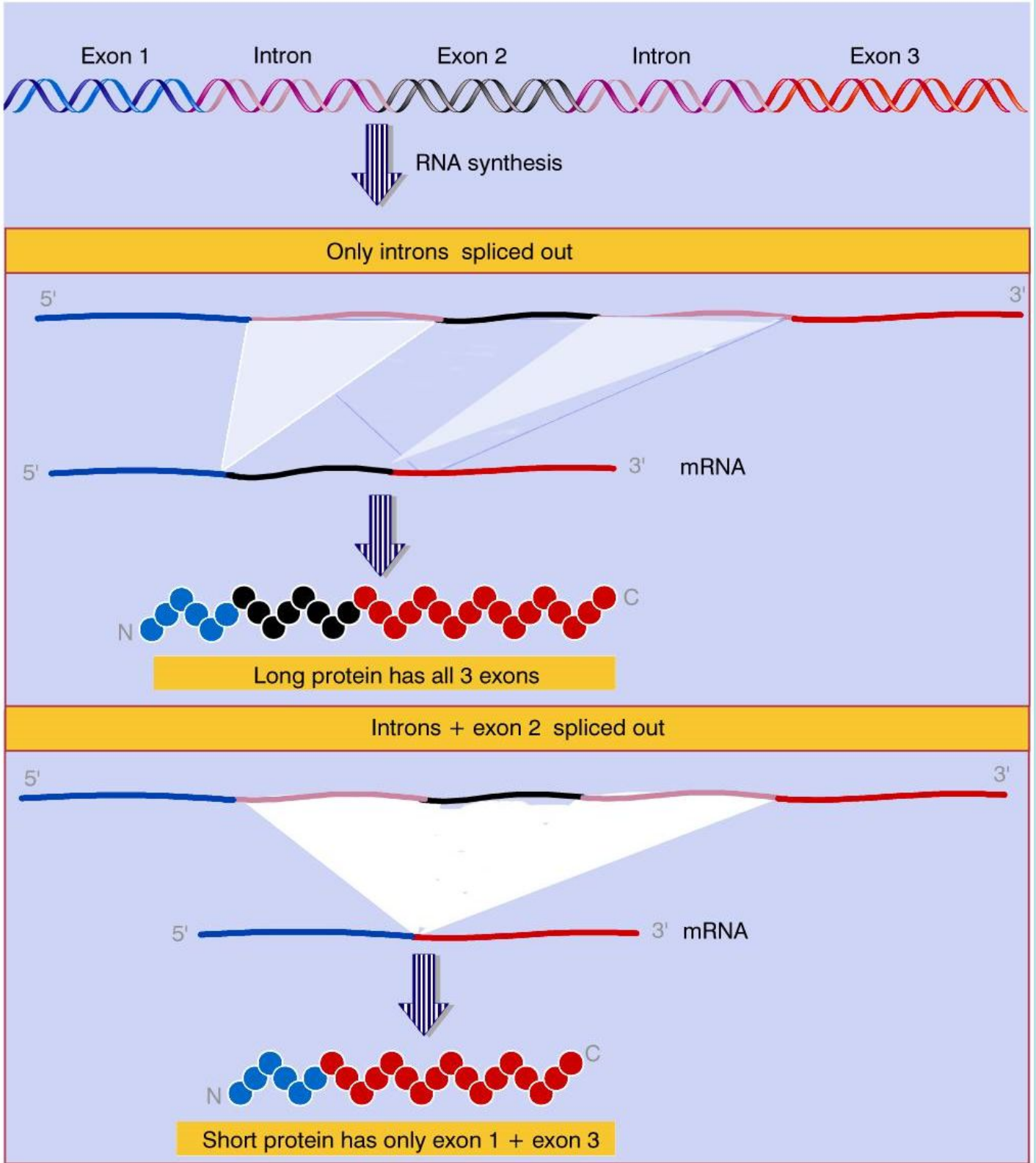


FIGURE 4.13 Two genes may overlap by reading the same DNA sequence in different frames.

PARTIALLY OVERLAPPING CODONS

RARE EVENT → for short sequences in some viral and mitochondrial genes

Figure 2.27 Alternative splicing uses the same pre-mRNA to generate mRNAs that have different combinations of exons.



TROPONIN T

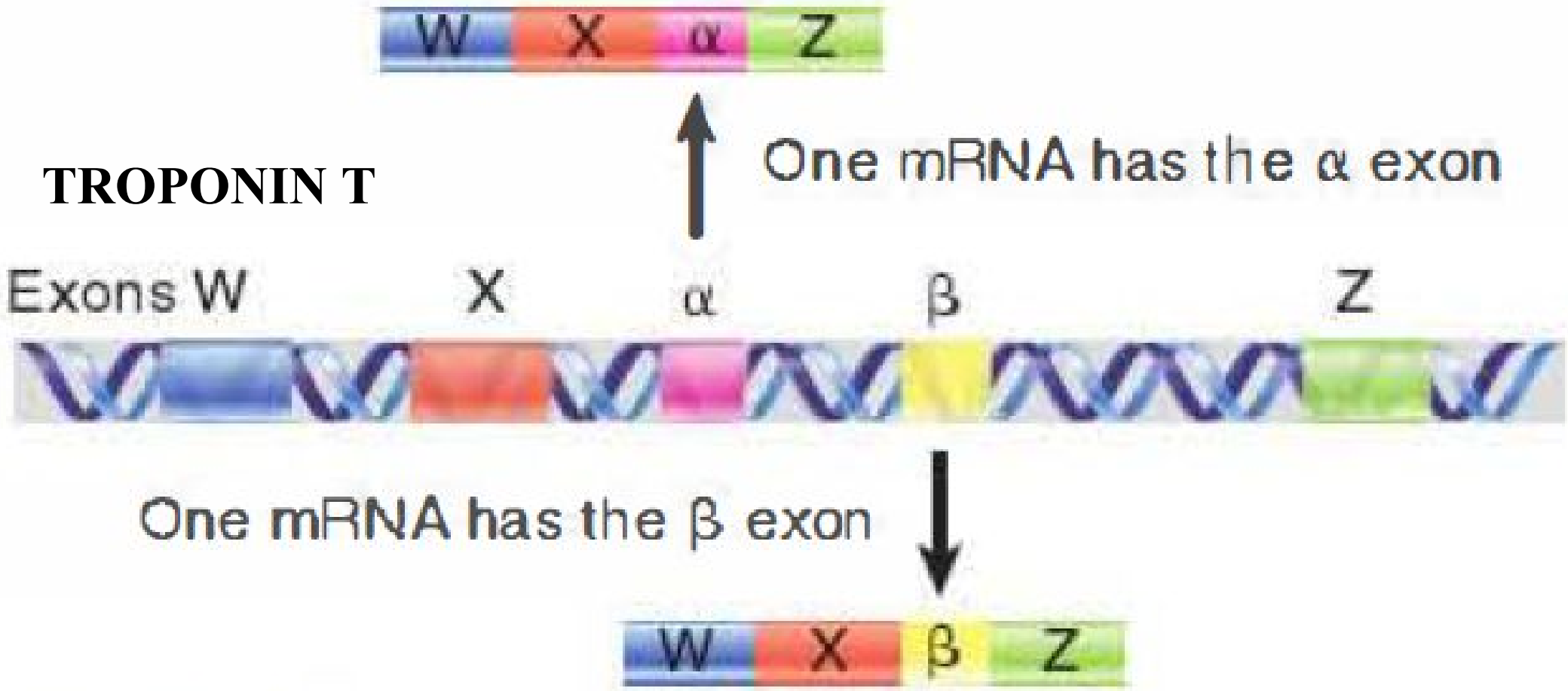
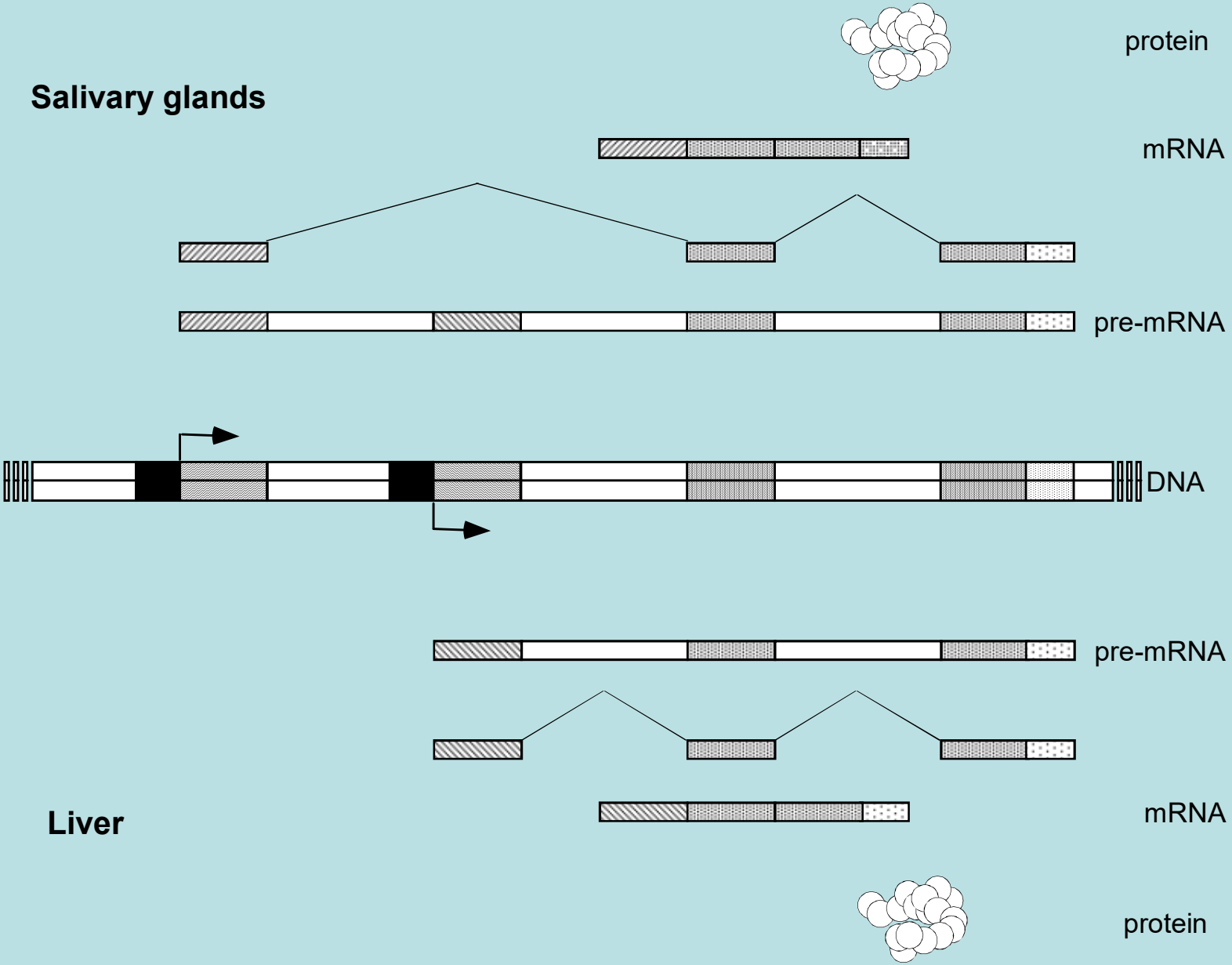
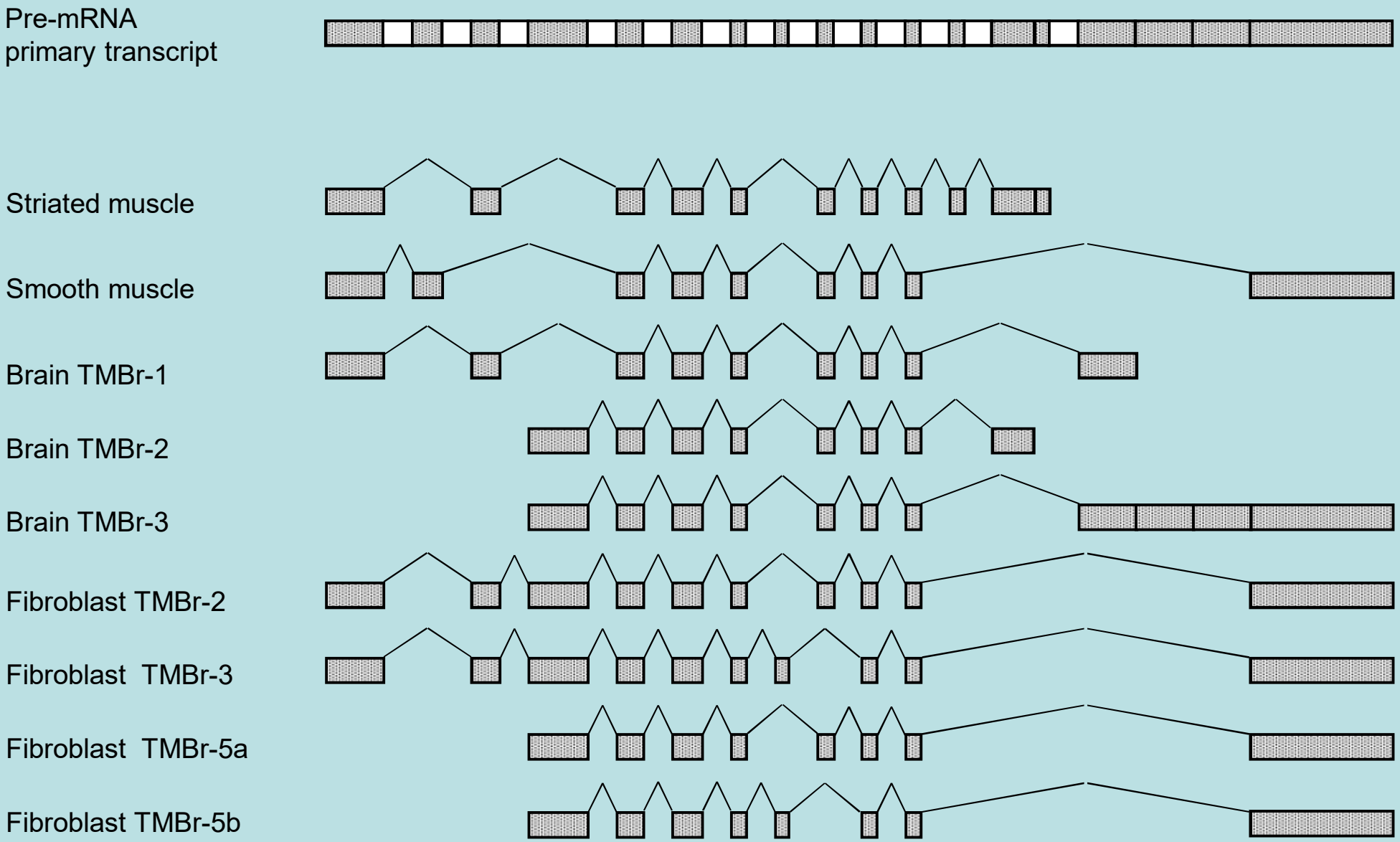


FIGURE 4.14 Alternative splicing generates the α and β variants of troponin T.

α-amylase gene



β-tropomyosin gene



Different mRNA isoforms can vary:

- in the coding region (the protein structure changes)

- in the non-translated region (changes in translation regulation and/or in messenger stability)