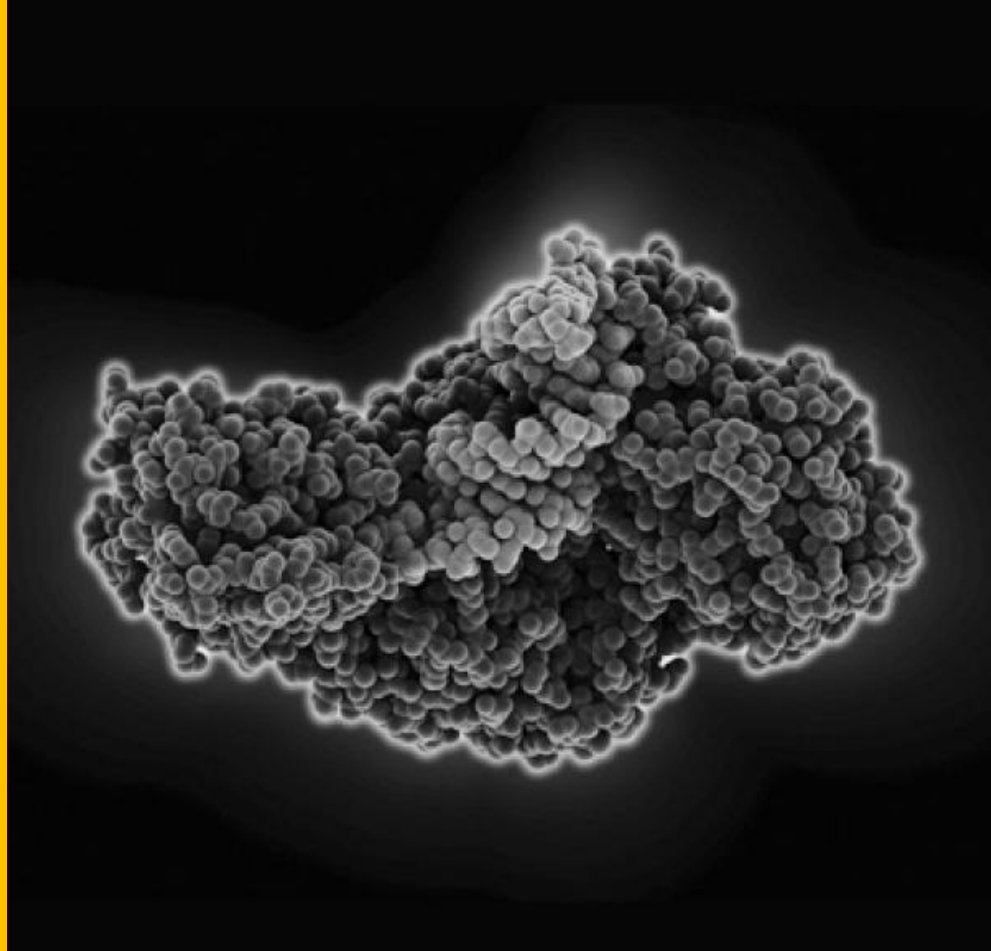
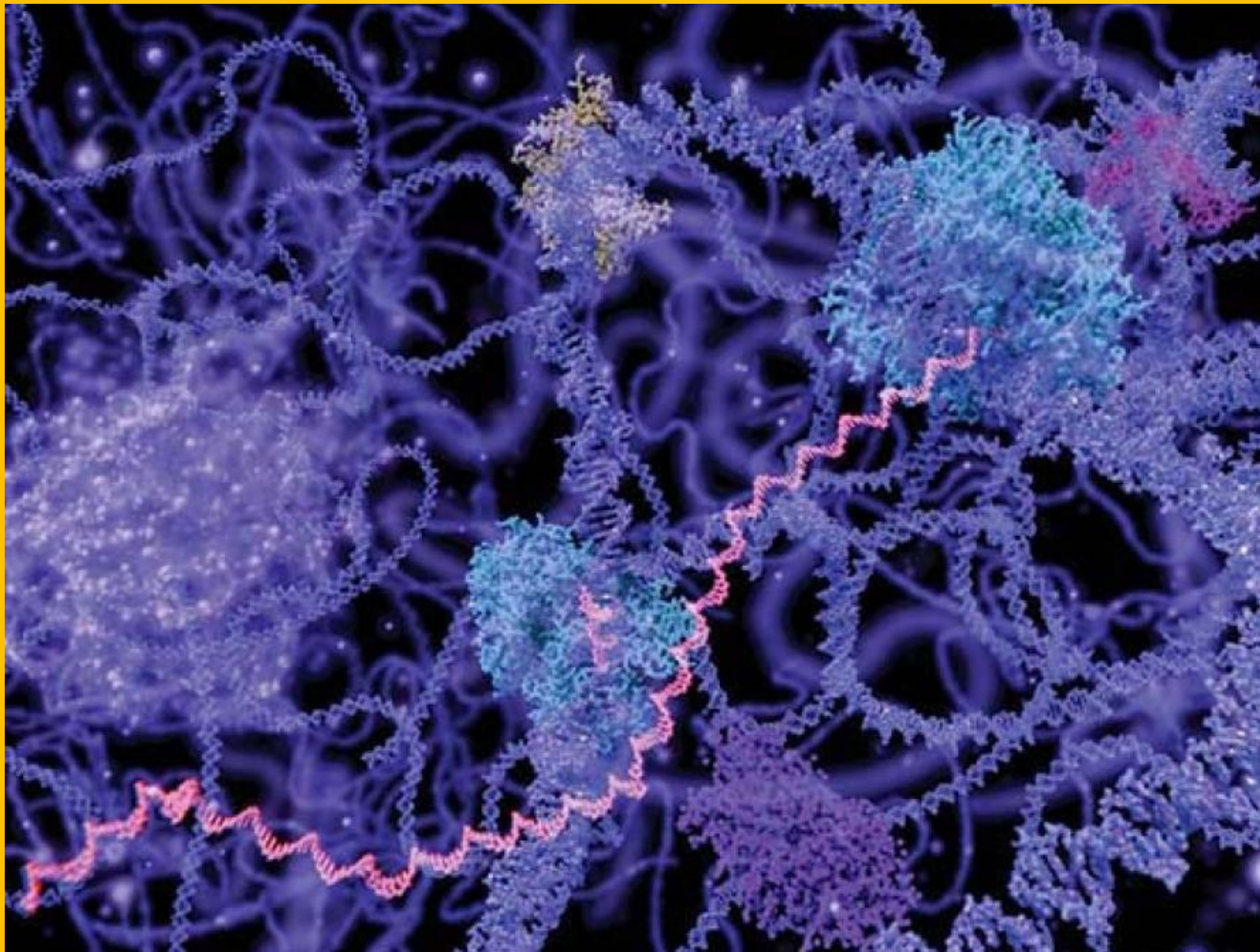


ΑΣΥΝΕΧΗ ΓΟΝΙΔΙΑ



Chapter 3: The Interrupted Gene



3.1 Introduction

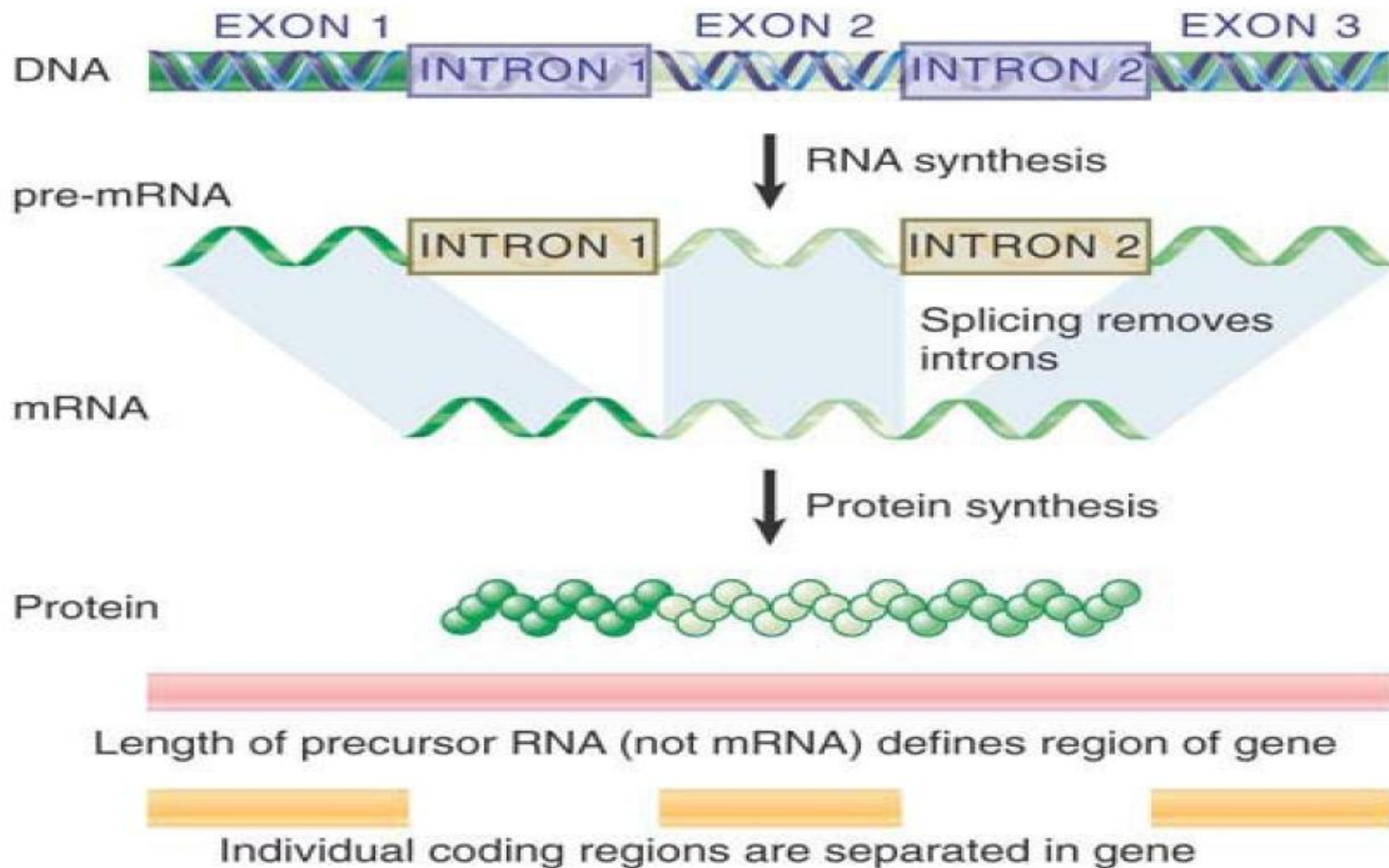
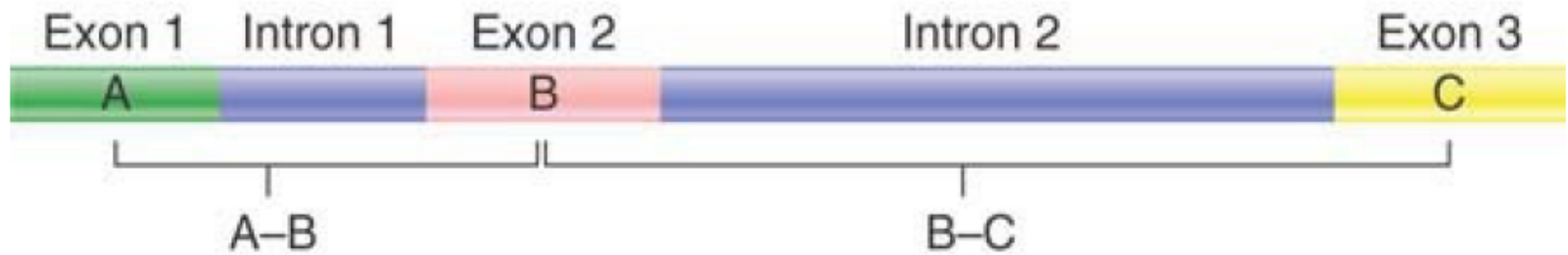


FIGURE 3.1 Interrupted genes are expressed via a precursor RNA. Introns are removed when the exons are spliced together. The mature mRNA has only the sequences of the exons.

3.2 An Interrupted Gene Has Exons and Introns

- Introns are removed by RNA splicing, which occurs in *cis* in individual RNA molecules.
- Mutations in exons can affect polypeptide sequence; mutations in introns can affect RNA processing and hence can influence the sequence and/or production of a polypeptide.

Genomic DNA



mRNA

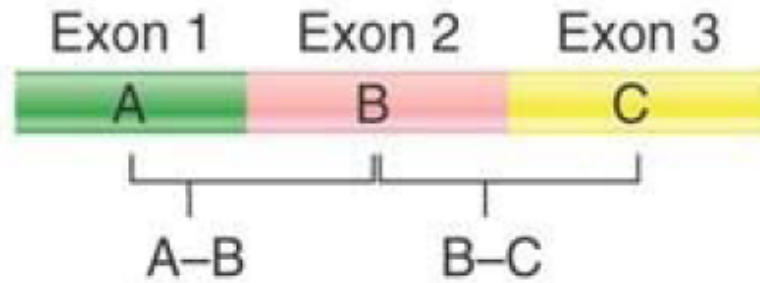


FIGURE 3.2 Exons remain in the same order in mRNA as in DNA, but distances along the gene do not correspond to distances along the mRNA or polypeptide products. The distance from A–B in the gene is smaller than the distance from B–C, but the distance from A–B in the mRNA (and polypeptide) is greater than the distance from B–C.

3.3 Exon and Intron Base Compositions Differ

- The four “rules” for DNA base composition are the first and second parity rules (both also known as Chargaff’s rules), the cluster rule, and the GC rule. Exons and introns can be distinguished on the basis of all rules except the first.
- The second parity rule suggests an extrusion of structured stem-loop segments from duplex DNA, which would be greater in introns.
- The rules relate to genomic characteristics, or “pressures,” that constitute the genome phenotype.

3.4 Organization of Interrupted Genes Can Be Conserved

- Introns can be detected when genes are compared with their RNA transcription products by sequencing.
- The positions of introns are usually conserved when homologous genes are compared between different organisms. The lengths of the corresponding introns can vary greatly.
- Introns usually do not encode proteins.

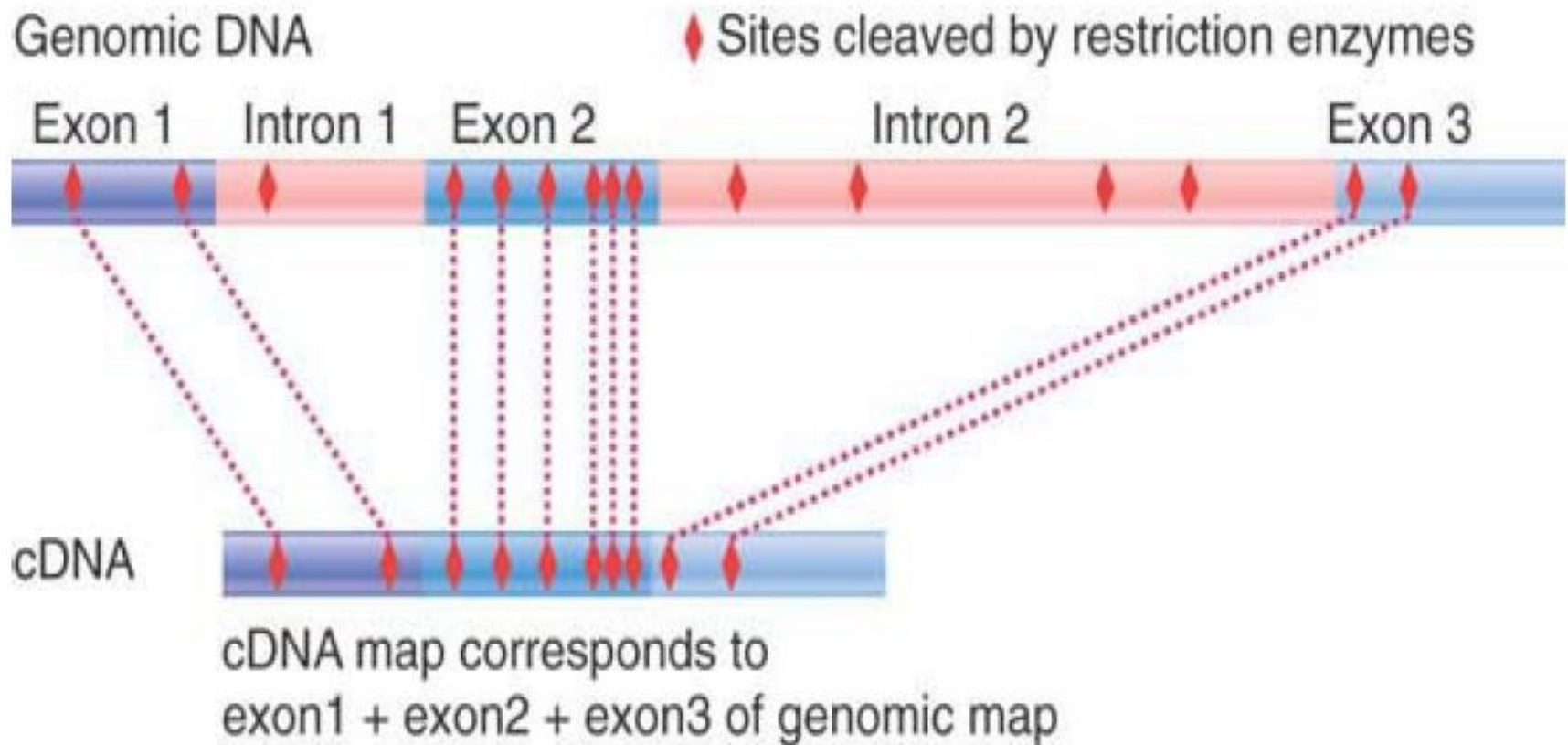


FIGURE 3.3 Comparison of the restriction maps of cDNA and genomic DNA for mouse β -globin shows that the gene has two introns that are not present in the cDNA. The exons can be aligned exactly between cDNA and the gene.

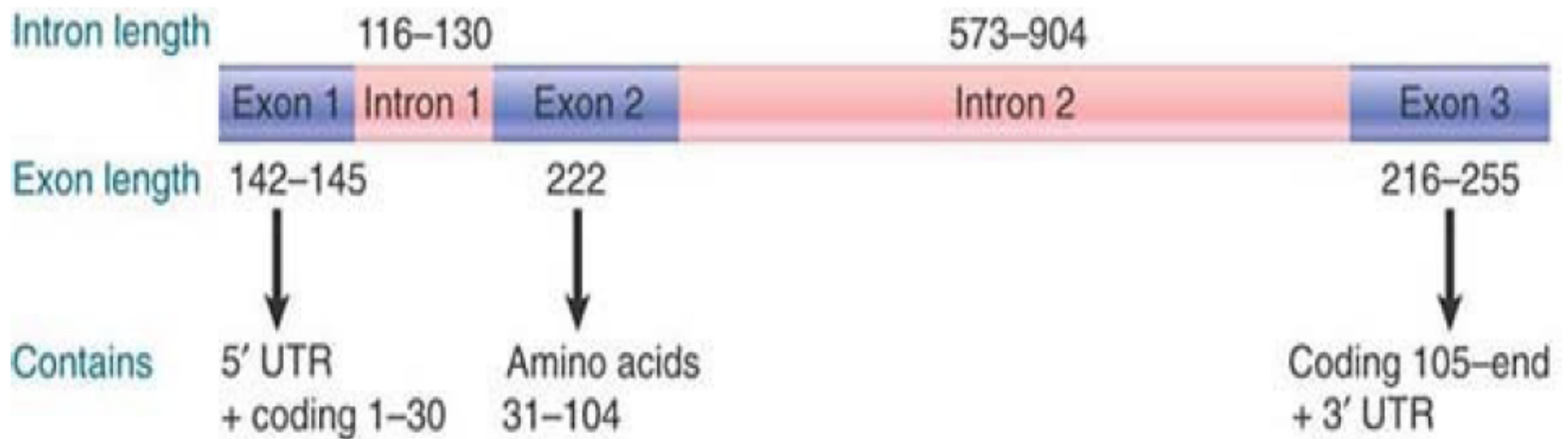


FIGURE 3.4 All functional globin genes have an interrupted structure with three exons. The lengths indicated in the figure apply to the mammalian β b-globin genes.

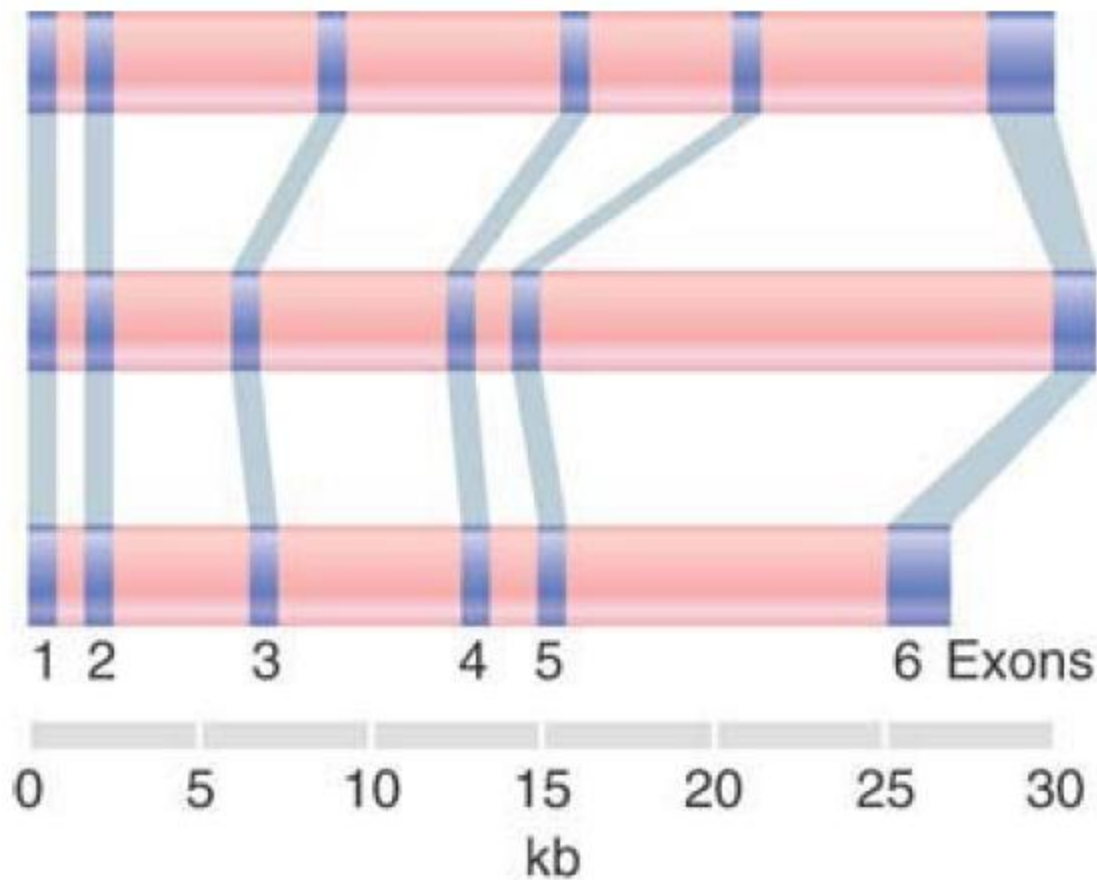


FIGURE 3.5 Mammalian genes for DHFR have the same relative organization of rather short exons and very long introns, but vary extensively in the lengths of introns.

3.5 Exon Sequences Under Negative Selection Are Conserved but Introns Vary

- Comparisons of related genes in different species show that the sequences of the corresponding exons are usually conserved but the sequences of the introns are much less similar.
- Introns evolve much more rapidly than exons because of the lack of selective pressure to produce a polypeptide with a useful sequence.

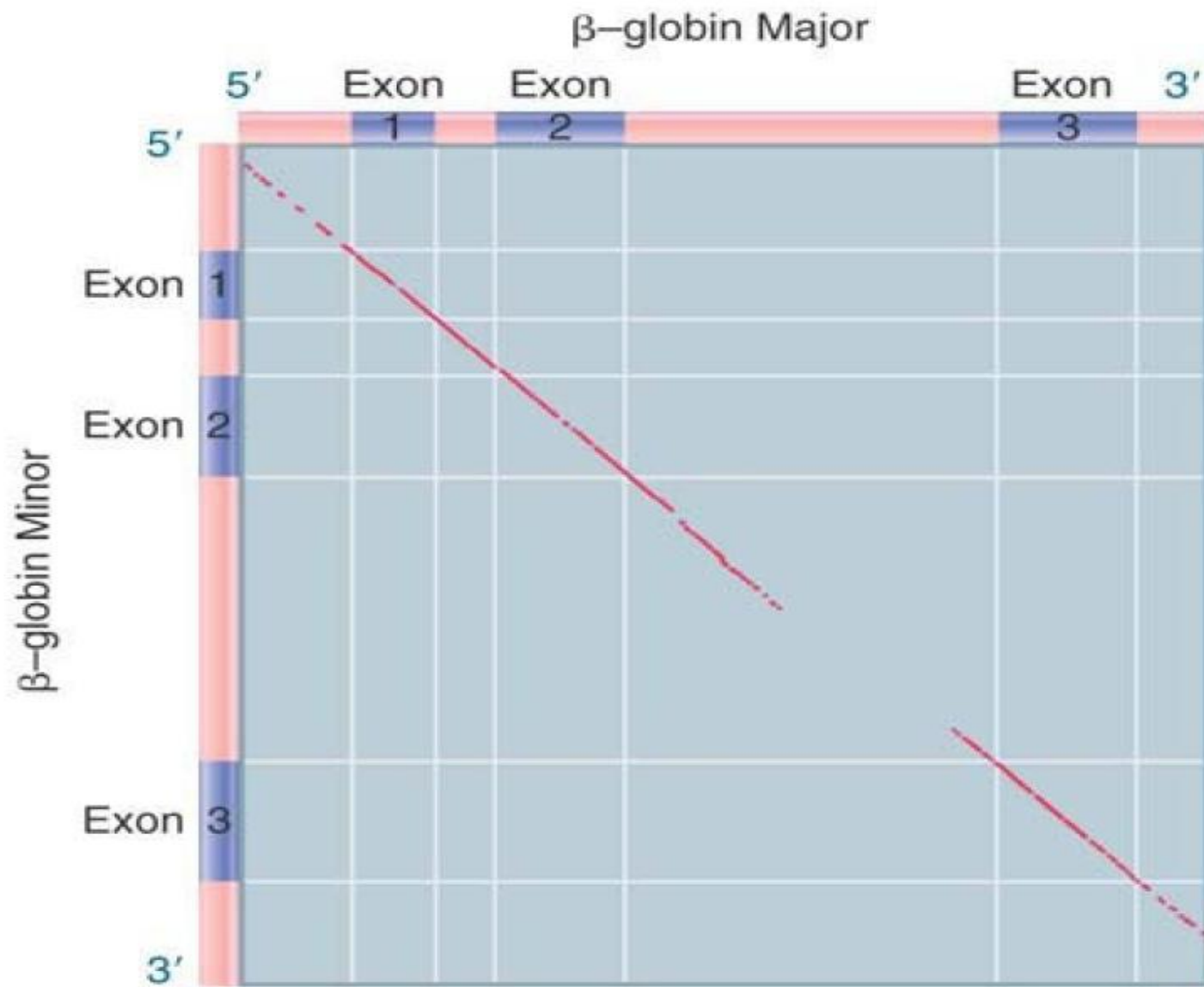


FIGURE 3.6 The sequences of the mouse $b\beta^{\text{maj-}}$ and $b\beta^{\text{min-}}$ -globin genes are closely related in coding regions but differ in the flanking UTRs and the long intron.

3.6 Exon Sequences Under Positive Selection Vary but Introns Are Conserved

- Under positive selection, an individual with an advantageous mutation survives (i.e., is able to produce more progeny that are fertile) relative to others without the mutation.
- Due to intrinsic genomic pressures, such as that which conserves the potential to extrude stem-loops from duplex DNA, introns evolve more slowly than exons that are under positive selection pressure.

3.7 Genes Show a Wide Distribution of Sizes Due Primarily to Intron Size and Number Variation

- Most genes are uninterrupted in *Saccharomyces cerevisiae* but are interrupted in multicellular eukaryotes.
- Exons are usually short, typically encoding fewer than 100 amino acids.
- Introns are short in unicellular/oligocellular eukaryotes but can be many kb in multicellular eukaryotes.
- The overall length of a gene is determined largely by its introns.

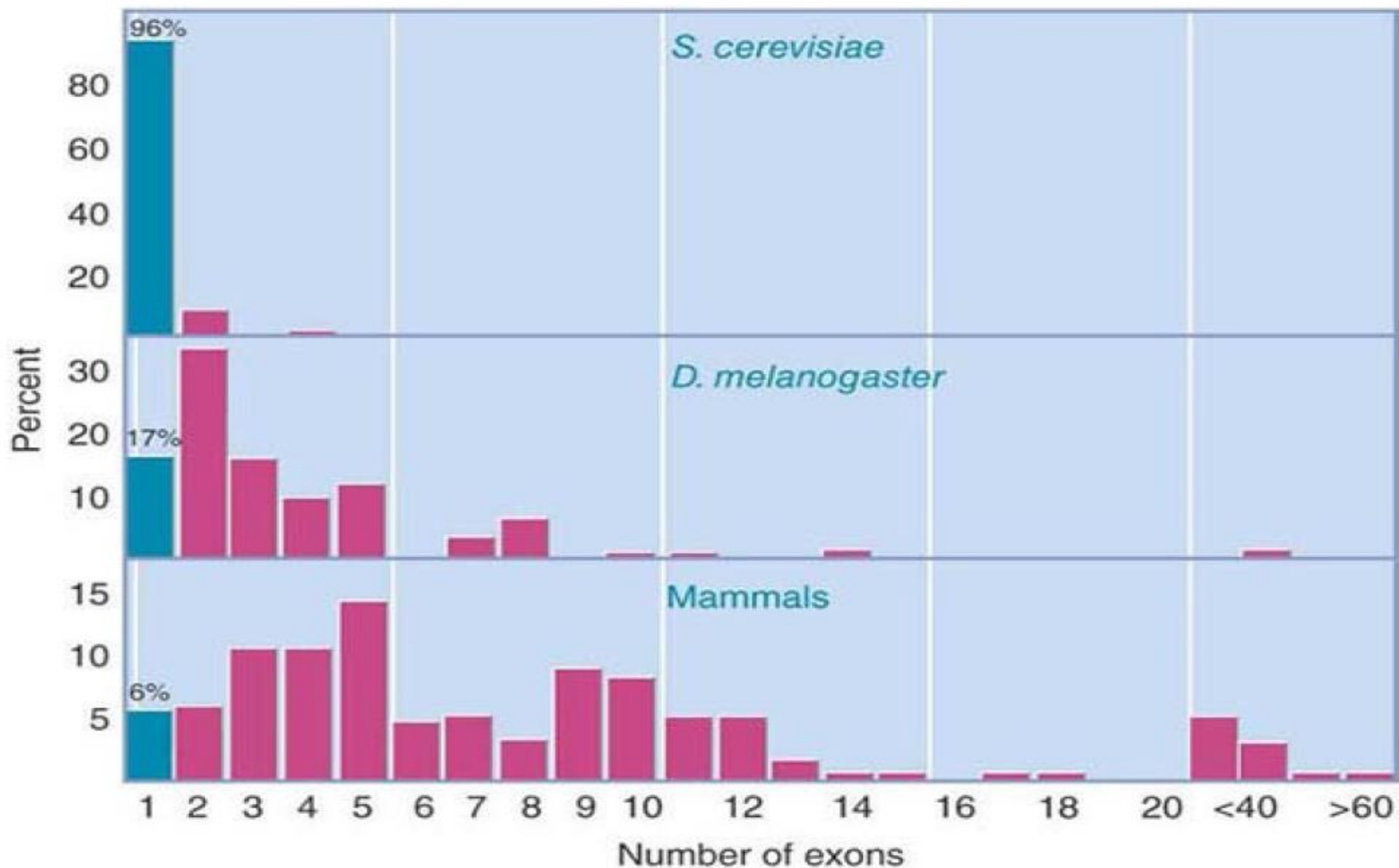


FIGURE 3.7 Most genes are uninterrupted in yeast, but most genes are interrupted in flies and mammals. (Uninterrupted genes have only one exon and are totaled in the leftmost column in blue.)

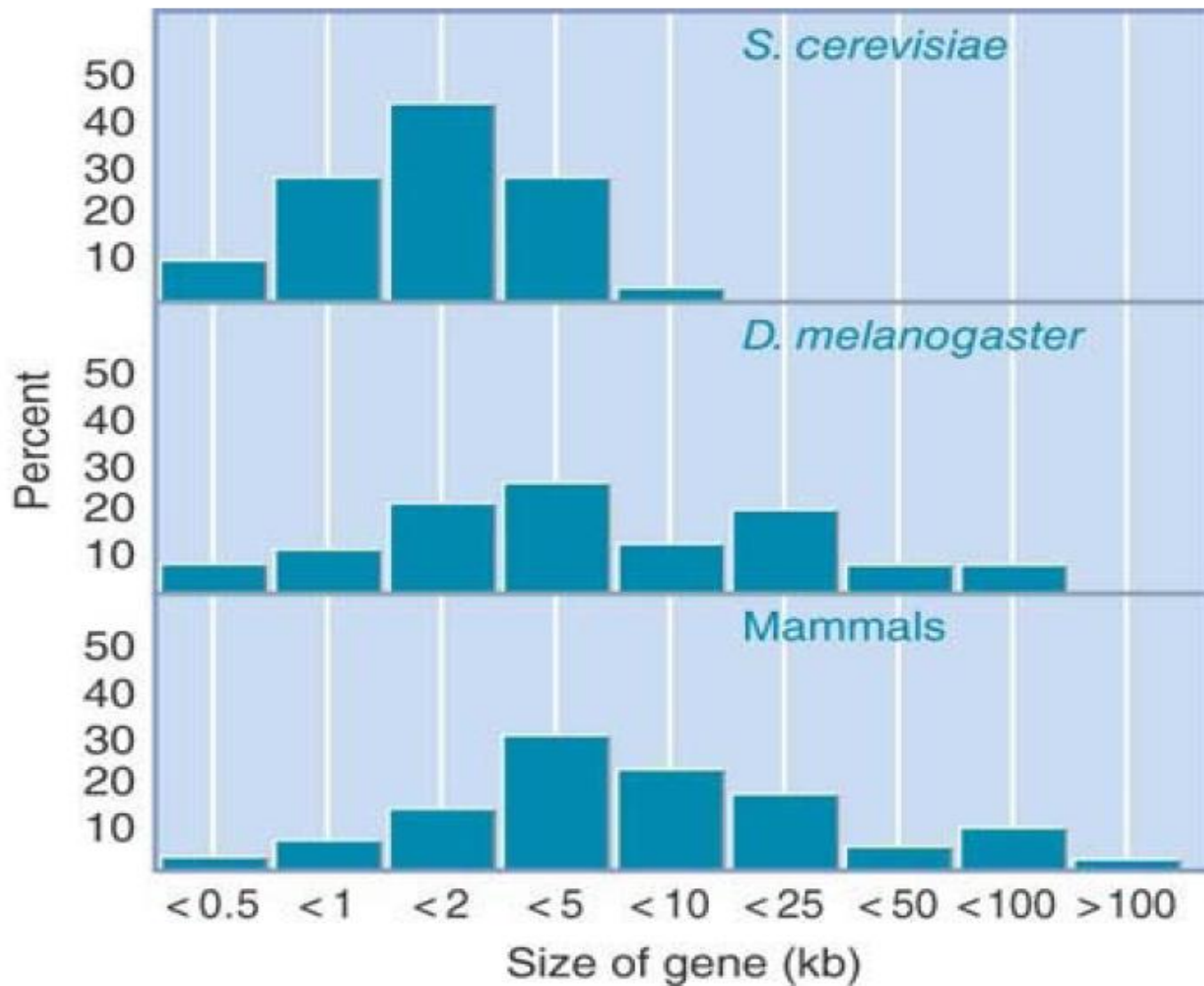


FIGURE 3.8 Yeast genes are short, but genes in flies and mammals have a dispersed bimodal distribution extending to very long sizes.

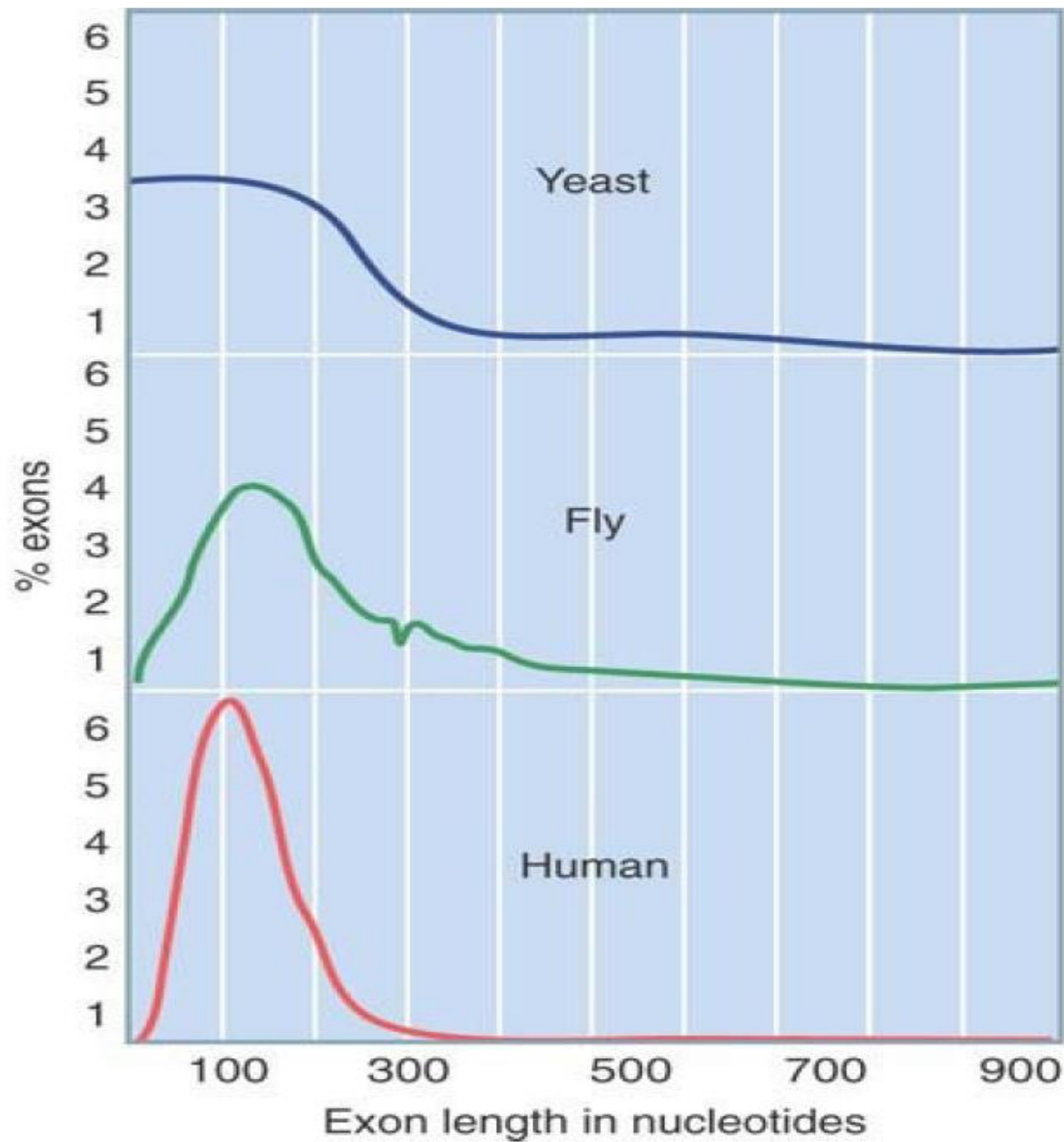


FIGURE 3.9 Exons encoding polypeptides are usually short.

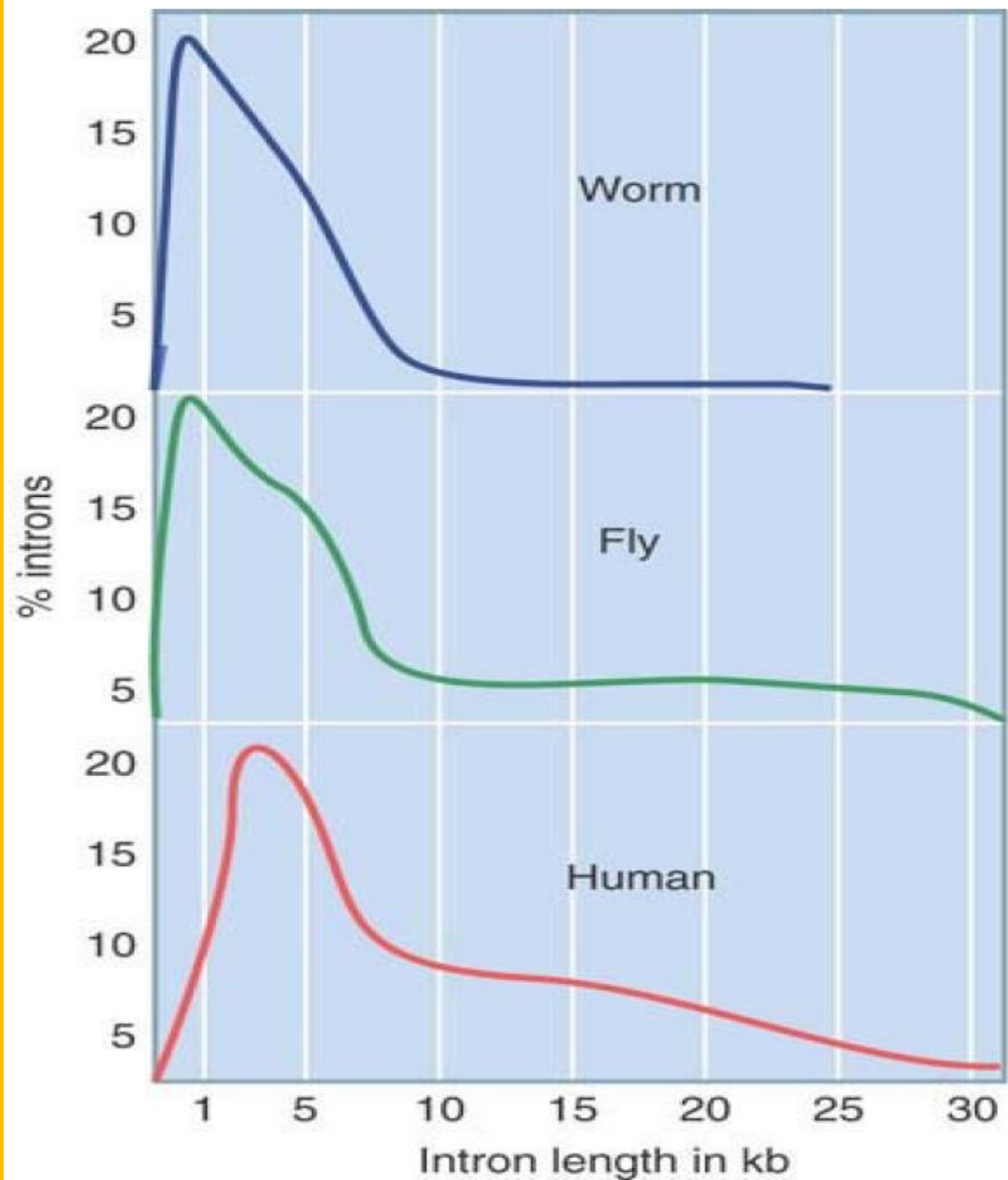


FIGURE 3.10 Introns range from very short to very long.

3.8 Some DNA Sequences Encode More Than One Polypeptide

- The use of alternative initiation or termination codons allows multiple variants of a polypeptide chain.
- Different polypeptides can be produced from the same sequence of DNA when the mRNA is read in different reading frames (as two overlapping genes).
- Otherwise identical polypeptides, differing by the presence or absence of certain regions, can be generated by differential (alternative) splicing. This can take the form of including or excluding individual exons, or of choosing between alternative exons.

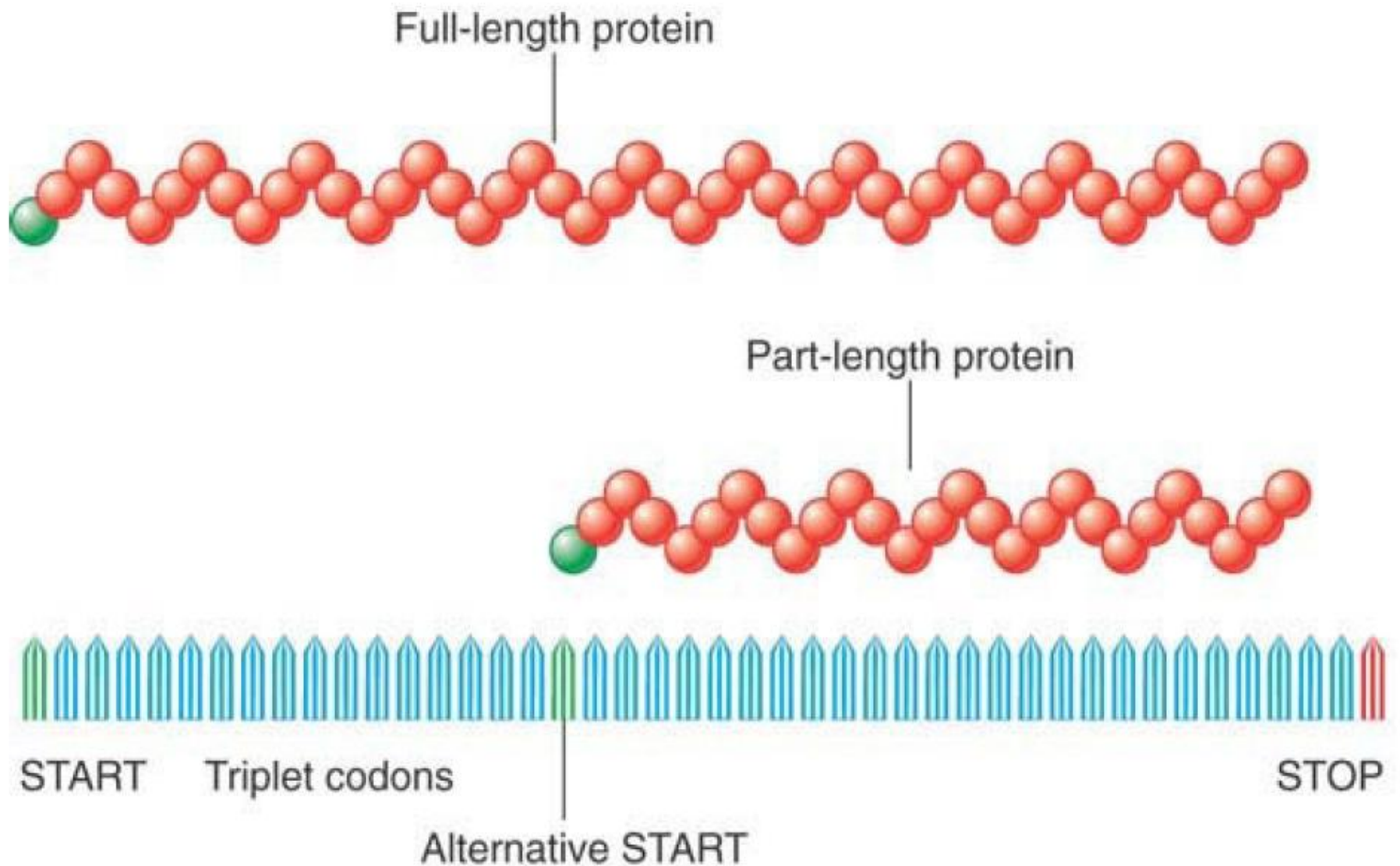


FIGURE 3.11 Two proteins can be generated from a single gene by starting (or terminating) expression at different points.

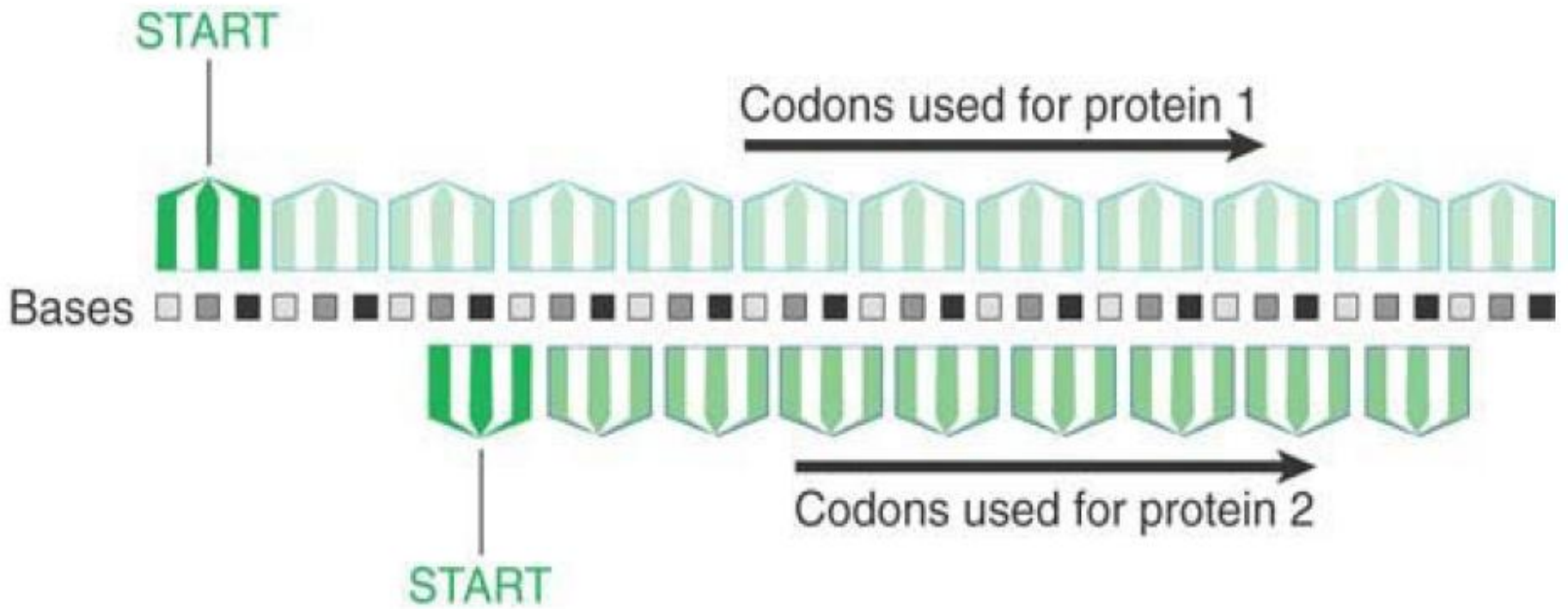


FIGURE 3.12 Two genes might overlap by reading the same DNA sequence in different frames.

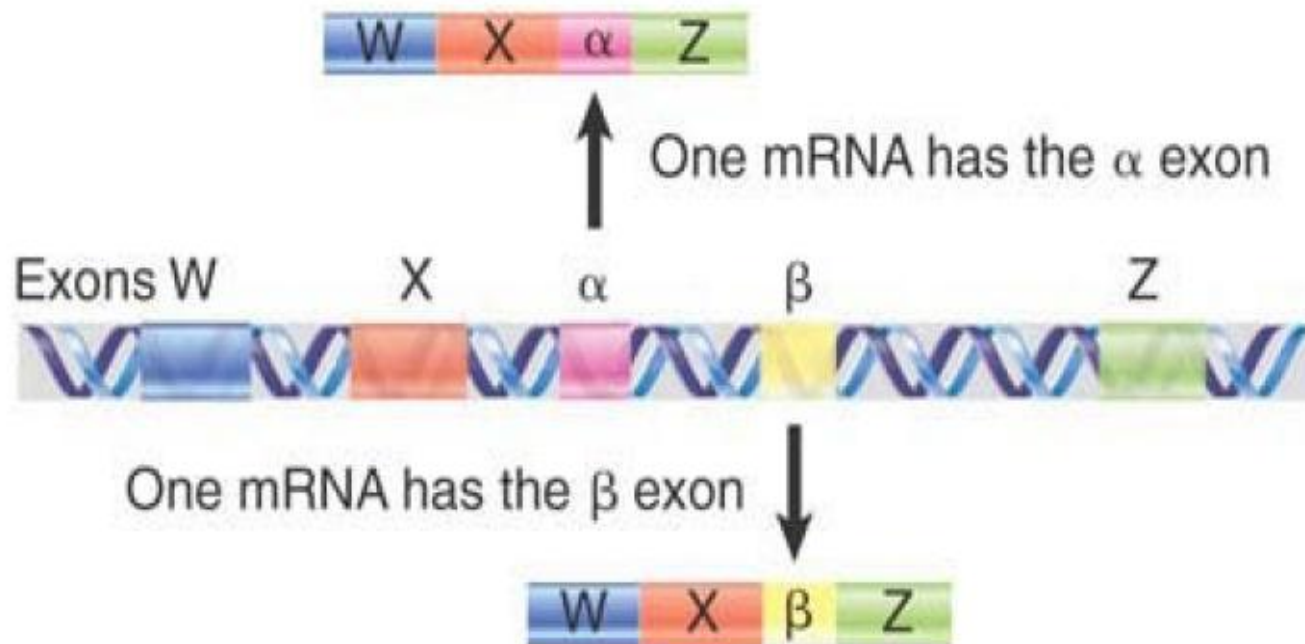


FIGURE 3.13 Alternative splicing generates the a and b variants of troponin T.

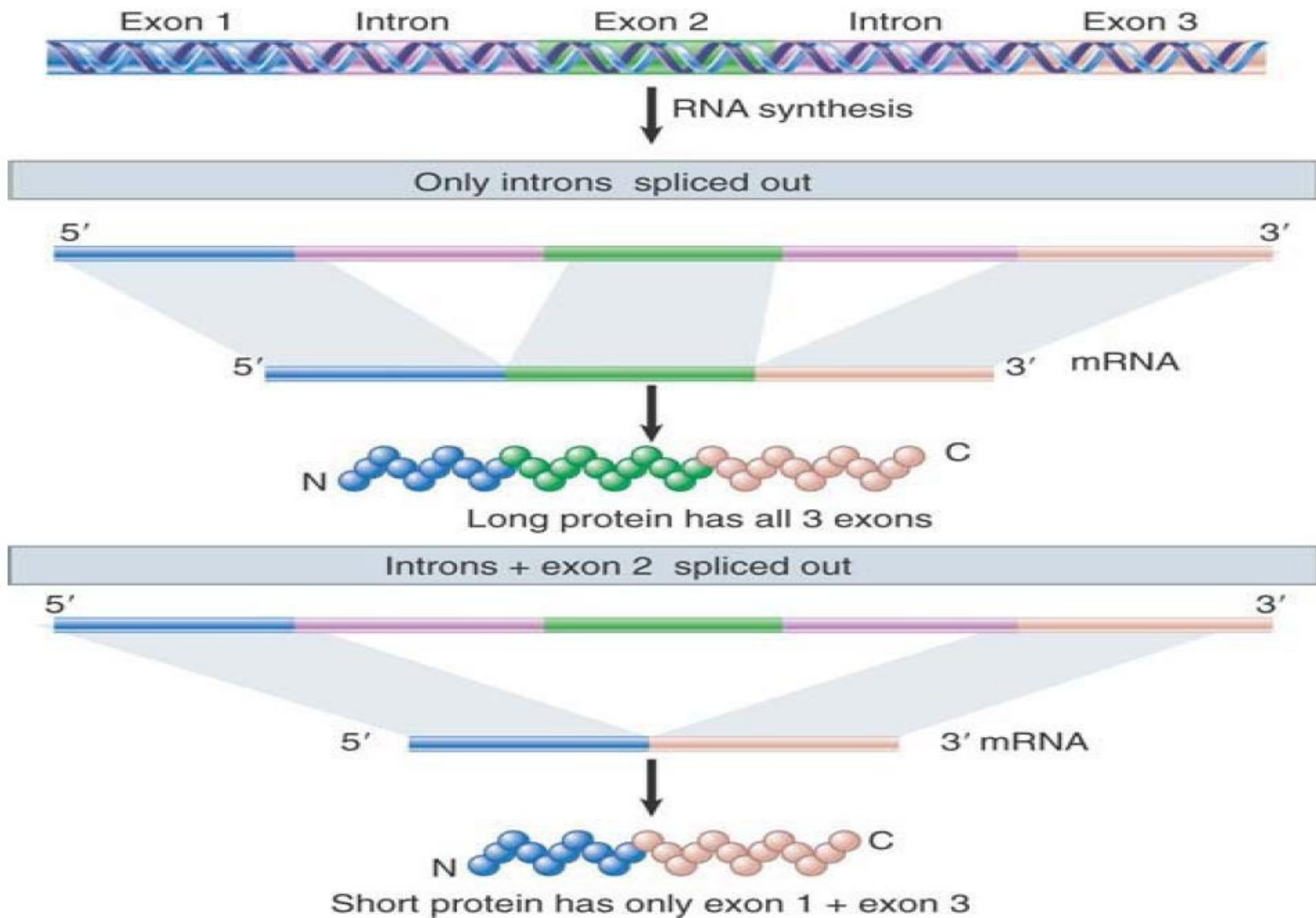


FIGURE 3.14 Alternative splicing uses the same pre-mRNA to generate mRNAs that have different combinations of exons.

3.9 Some Exons Correspond to Protein Functional Domains

- Proteins can consist of independent functional modules, the boundaries of which, in some cases, correspond to those of exons.
- The exons of some genes appear homologous to the exons of others, suggesting a common exon ancestry.

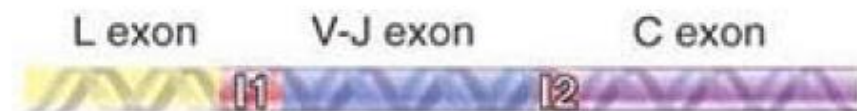


FIGURE 3.15 Immunoglobulin light chains and heavy chains are encoded by genes whose structures (in their expressed forms) correspond to the distinct domains in the protein. Each protein domain corresponds to an exon; introns are numbered I1 to I5.

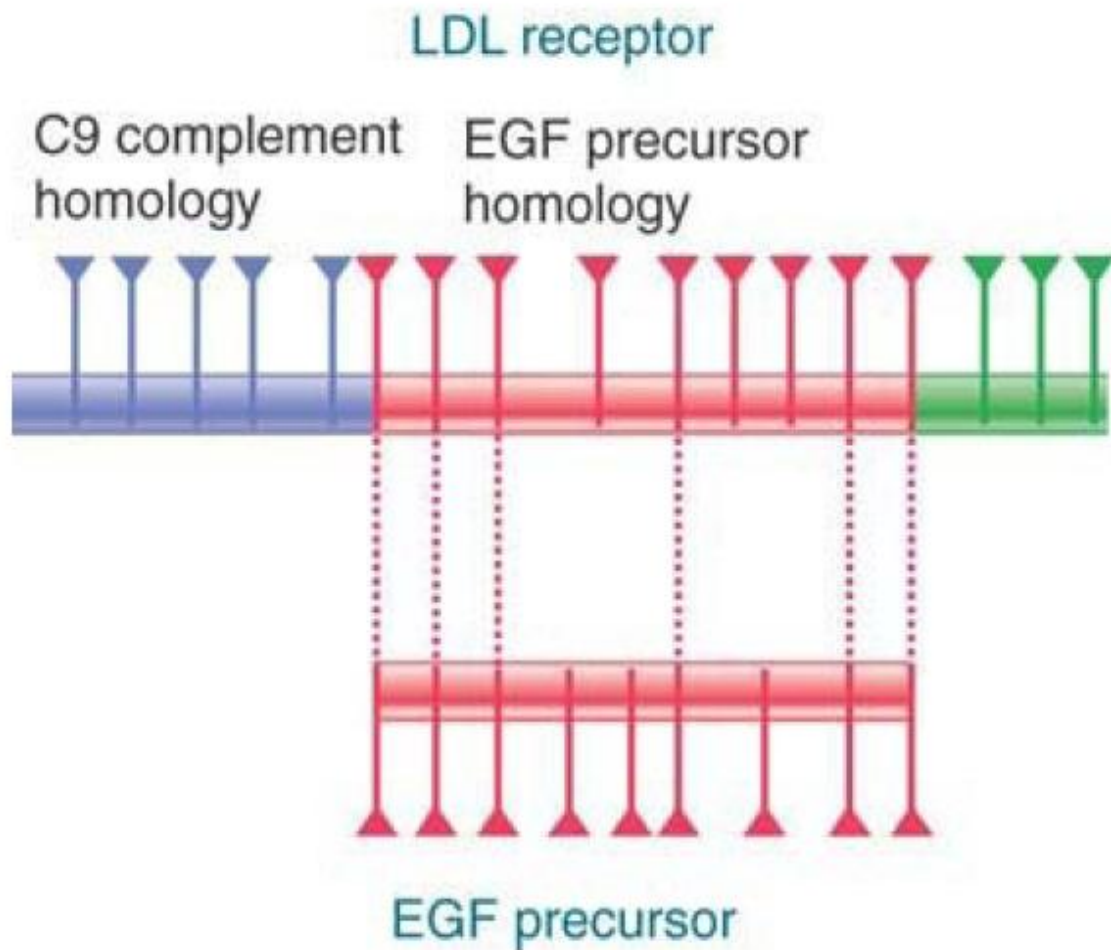


FIGURE 3.16 The LDL receptor gene consists of 18 exons, some of which are related to EGF precursor exons and some of which are related to the C9 blood complement gene. Triangles mark the positions of introns.

3.10 Members of a Gene Family Have a Common Organization

- A set of homologous genes should share common features that preceded their evolutionary separation.
- All globin genes have a common form of organization with three exons and two introns, suggesting that they are descended from a single ancestral gene.
- Intron positions in the actin gene family are highly variable, which suggests that introns do not separate functional domains.

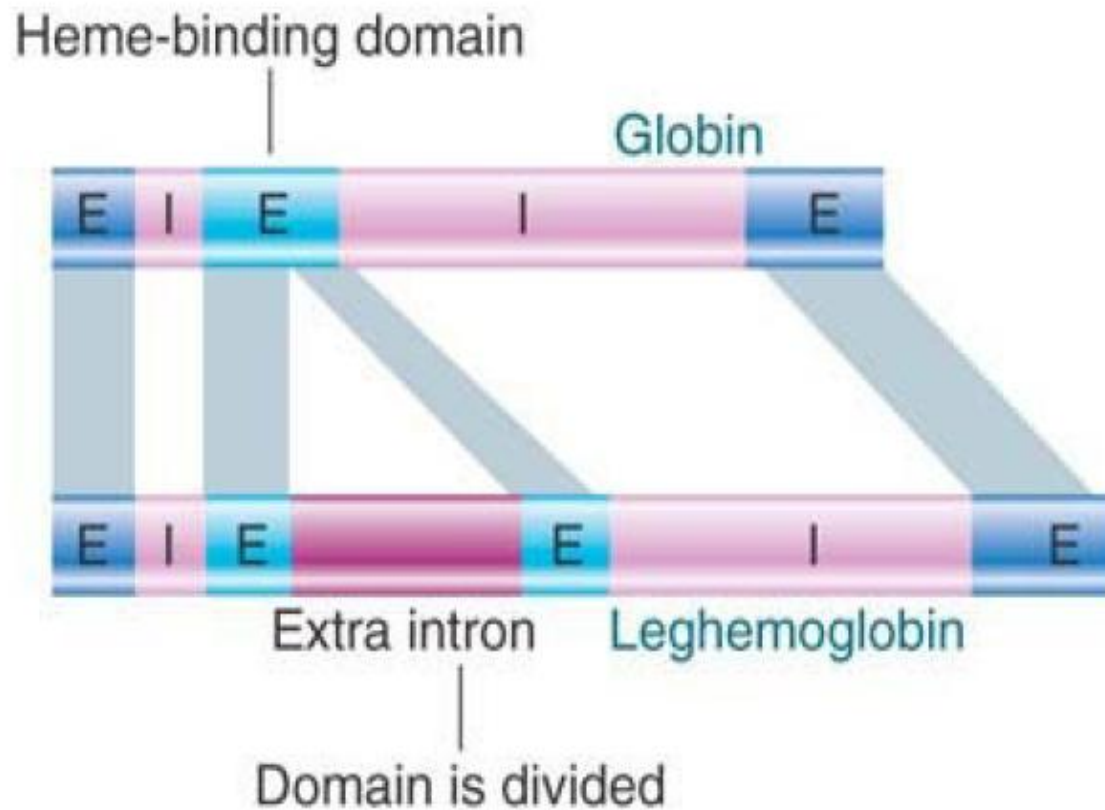


FIGURE 3.17 The exon structure of globin genes corresponds to protein function, but leghemoglobin has an extra intron in the central domain.

Common insulin gene (chicken and rat)

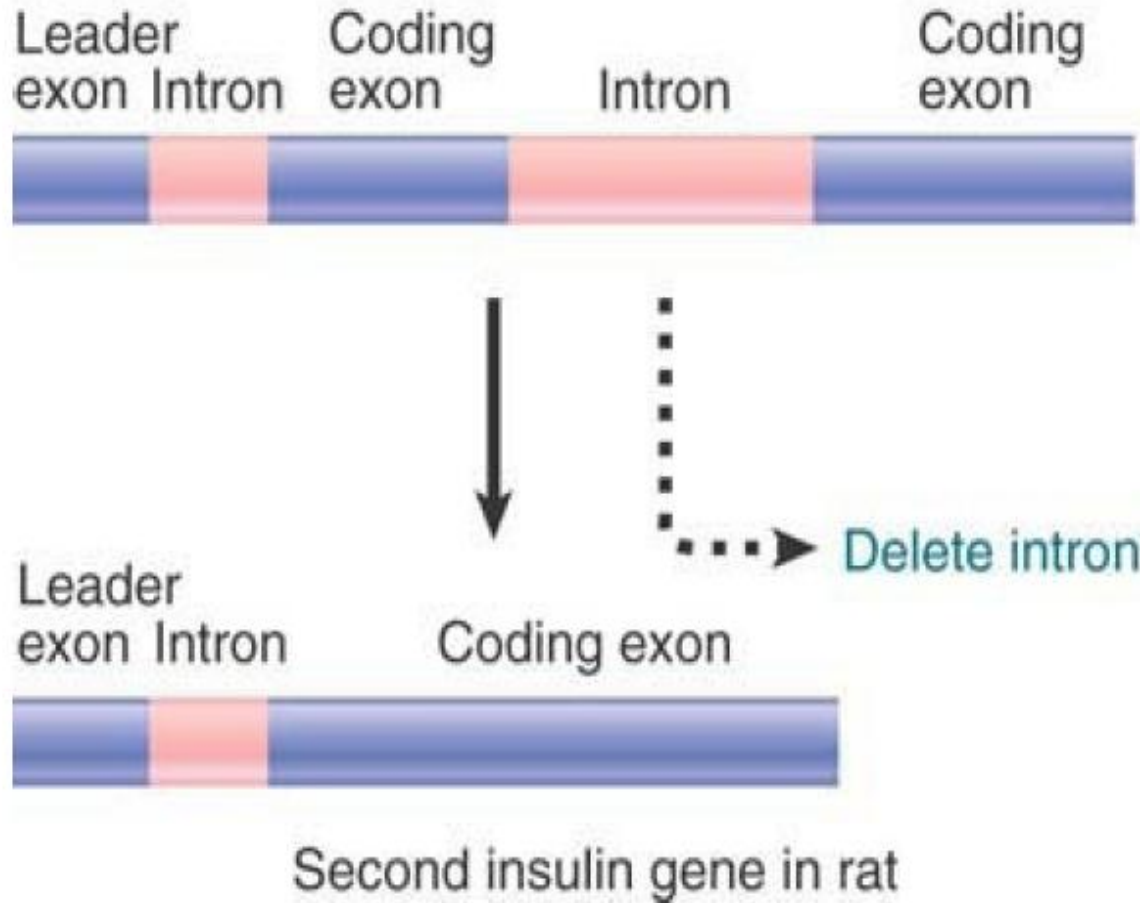


FIGURE 3.18 The rat insulin gene with one intron evolved by loss of an intron from an ancestor with two introns.

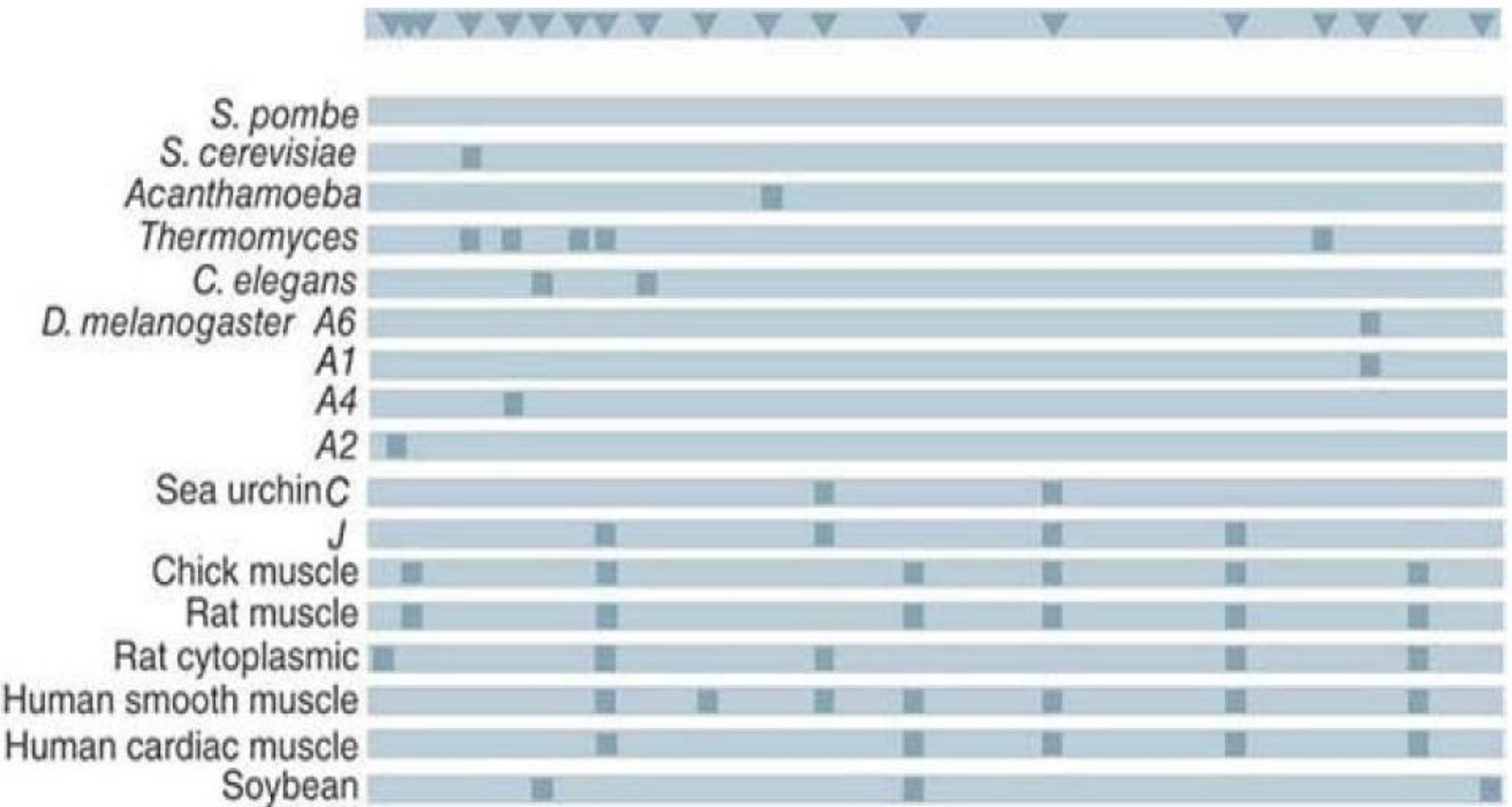


FIGURE 3.19 Actin genes vary widely in their organization. The sites of introns are indicated by dark boxes. The bar at the top summarizes all the intron positions among the different orthologs.

3.11 There Are Many Forms of Information in DNA

- Genetic information includes not only that related to characters corresponding to the conventional phenotype but also that related to characters (pressures) corresponding to the genome “phenotype.”
- In certain contexts, the definition of the gene can be seen as reversed from “one gene–one protein” to “one protein–one gene.”
- Positional information might be important in development.
- Sequences transferred “horizontally” from other species to the germ line could land in introns or intergenic DNA and then transfer “vertically” through the generations. Some of these sequences might be involved in intracellular non-self-recognition.