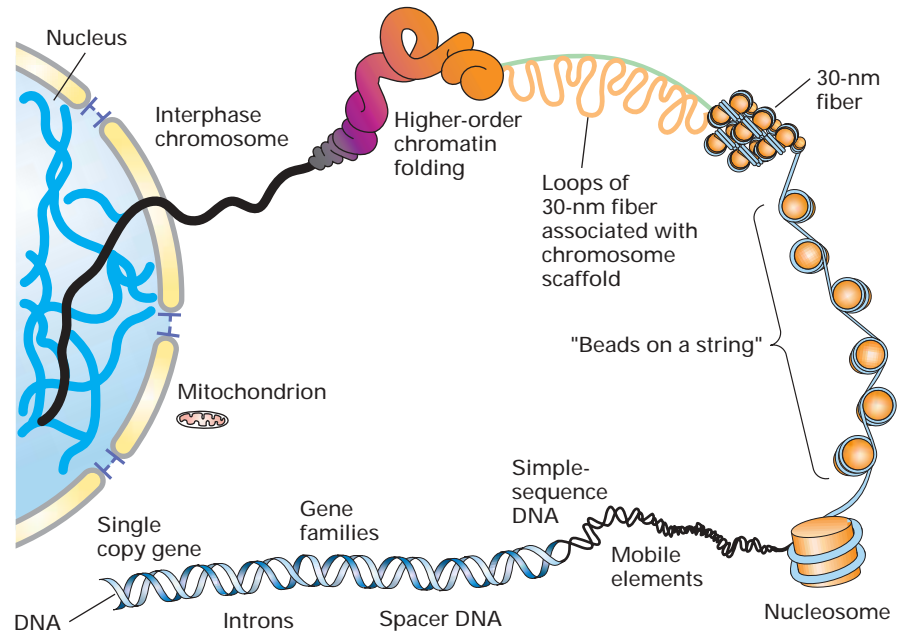


► **FIGURE 10-1 Overview of the structure of genes and chromosomes.**

DNA of higher eukaryotes consists of unique and repeated sequences. Only ~5% of human DNA encodes proteins and functional RNAs and the regulatory sequences that control their expression; the remainder is merely spacer DNA between genes and introns within genes. Much of this DNA, ~50% in humans, is derived from mobile DNA elements, genetic symbiots that have contributed to the evolution of contemporary genomes. Each chromosome consists of a single, long molecule of DNA up to ~280 Mb in humans, organized into increasing levels of condensation by the histone and nonhistone proteins with which it is intricately complexed. Much smaller DNA molecules are localized in mitochondria and chloroplasts.



chloroplasts also contain DNA, probably evolutionary remnants of their origins, that encode essential components of these vital organelles.

In this chapter we first present a molecular definition of genes and the complexities that arise in higher organisms from the processing of mRNA precursors into alternatively spliced mRNAs. Next we discuss the main classes of eukaryotic DNA and the special properties of mobile DNA. We also consider the packaging of DNA and proteins into compact complexes, the large-scale structure of chromosomes, and the functional elements required for chromosome duplication and segregation. In the final section, we consider organelle DNA and how it differs from nuclear DNA. Figure 10-1 provides an overview of these interrelated subjects.

10.1 Molecular Definition of a Gene

In molecular terms, a gene commonly is defined as *the entire nucleic acid sequence that is necessary for the synthesis of a functional gene product (polypeptide or RNA)*. According to this definition, a gene includes more than the nucleotides encoding the amino acid sequence of a protein, referred to as the *coding region*. A gene also includes all the DNA sequences required for synthesis of a particular RNA transcript. In eukaryotic genes, transcription-control regions known as **enhancers** can lie 50 kb or more from the coding region. Other critical noncoding regions in eukaryotic genes are the sequences that specify 3' cleavage and polyadenylation, known as *poly(A) sites*, and splicing of primary RNA transcripts, known as *splice sites* (see Figure 4-14). Mutations in these RNA-processing signals prevent expression of a functional mRNA and thus of the encoded polypeptide.

Although most genes are transcribed into mRNAs, which encode proteins, clearly some DNA sequences are transcribed into RNAs that do not encode proteins (e.g., tRNAs and rRNAs). However, because the DNA that encodes tRNAs and rRNAs can cause specific phenotypes when it is mutated, these DNA regions generally are referred to as tRNA and rRNA *genes*, even though the final products of these genes are RNA molecules and not proteins. Many other RNA molecules described in later chapters also are transcribed from non-protein-coding genes.

Most Eukaryotic Genes Produce Monocistronic mRNAs and Contain Lengthy Introns

As discussed in Chapter 4, many bacterial mRNAs are *polycistronic*; that is, a single mRNA molecule (e.g., the mRNA encoded by the *trp* operon) includes the coding region for several proteins that function together in a biological process. In contrast, most eukaryotic mRNAs are *monocistronic*; that is, each mRNA molecule encodes a single protein. This difference between polycistronic and monocistronic mRNAs correlates with a fundamental difference in their translation.

Within a bacterial polycistronic mRNA a ribosome-binding site is located near the start site for each of the protein-coding regions, or cistrons, in the mRNA. Translation initiation can begin at any of these multiple internal sites, producing multiple proteins (see Figure 4-12a). In most eukaryotic mRNAs, however, the 5'-cap structure directs ribosome binding, and translation begins at the closest AUG start codon (see Figure 4-12b). As a result, translation begins only at this site. In many cases, the primary transcripts of eukaryotic protein-coding genes are processed into a single type of