

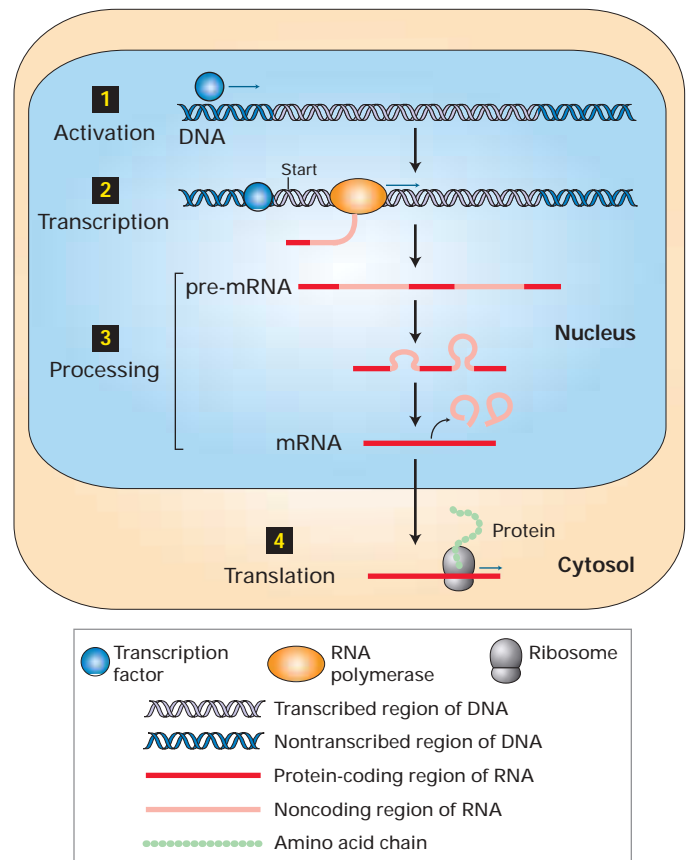
even if the solution also contains many other DNA strands that do not match.

The genetic information carried by DNA resides in its sequence, the linear order of nucleotides along a strand. The information-bearing portion of DNA is divided into discrete functional units, the **genes**, which typically are 5000 to 100,000 nucleotides long. Most bacteria have a few thousand genes; humans, about 40,000. The genes that carry instructions for making proteins commonly contain two parts: a *coding region* that specifies the amino acid sequence of a protein and a *regulatory region* that controls when and in which cells the protein is made.

Cells use two processes in series to convert the coded information in DNA into proteins (Figure 1-11). In the first, called **transcription**, the coding region of a gene is copied into a single-stranded **ribonucleic acid (RNA)** version of the double-stranded DNA. A large enzyme, **RNA polymerase**, catalyzes the linkage of nucleotides into a RNA chain using DNA as a template. In eukaryotic cells, the initial RNA product is processed into a smaller **messenger RNA (mRNA)** molecule, which moves to the cytoplasm. Here the **ribosome**, an enormously complex molecular machine composed of both RNA and protein, carries out the second process, called **translation**. During translation, the ribosome assembles and links together amino acids in the precise order dictated by the mRNA sequence according to the nearly universal **genetic code**. We examine the cell components that carry out transcription and translation in detail in Chapter 4.

All organisms have ways to control when and where their genes can be transcribed. For instance, nearly all the cells in our bodies contain the full set of human genes, but in each cell type only some of these genes are active, or turned on, and used to make proteins. That's why liver cells produce some proteins that are not produced by kidney cells, and vice versa. Moreover, many cells can respond to external signals or changes in external conditions by turning specific genes on or off, thereby adapting their repertoire of proteins to meet current needs. Such control of gene activity depends on DNA-binding proteins called **transcription factors**, which bind to DNA and act as switches, either activating or repressing transcription of particular genes (Chapter 11).

Transcription factors are shaped so precisely that they are able to bind preferentially to the regulatory regions of just a few genes out of the thousands present in a cell's DNA. Typically a DNA-binding protein will recognize short DNA sequences about 6–12 base pairs long. A segment of DNA containing 10 base pairs can have 4^{10} possible sequences (1,048,576) since each position can be any of four nucleotides. Only a few copies of each such sequence will occur in the DNA of a cell, assuring the specificity of gene activation and repression. Multiple copies of one type of transcription factor can coordinately regulate a set of genes if binding sites for that factor exist near each gene in the set. Transcription factors often work as multiprotein complexes, with more than one protein contributing its own DNA-binding specificity to selecting the regulated genes. In complex organisms,



▲ **FIGURE 1-11 The coded information in DNA is converted into the amino acid sequences of proteins by a multistep process.** Step **1**: Transcription factors bind to the regulatory regions of the specific genes they control and activate them. Step **2**: Following assembly of a multiprotein initiation complex bound to the DNA, RNA polymerase begins transcription of an activated gene at a specific location, the start site. The polymerase moves along the DNA linking nucleotides into a single-stranded pre-mRNA transcript using one of the DNA strands as a template. Step **3**: The transcript is processed to remove noncoding sequences. Step **4**: In a eukaryotic cell, the mature messenger RNA (mRNA) moves to the cytoplasm, where it is bound by ribosomes that read its sequence and assemble a protein by chemically linking amino acids into a linear chain.

hundreds of different transcription factors are employed to form an exquisite control system that activates the right genes in the right cells at the right times.

The Genome Is Packaged into Chromosomes and Replicated During Cell Division

Most of the DNA in eukaryotic cells is located in the nucleus, extensively folded into the familiar structures we know as **chromosomes** (Chapter 10). Each chromosome contains a single linear DNA molecule associated with certain proteins. In prokaryotic cells, most or all of the genetic information resides