Figure 6–1 (opposite page) Schematic depiction of a portion of chromosome 2 from the genome of the fruit fly Drosophila melanogaster. This figure represents approximately 3% of the total Drosophila genome, arranged as six contiguous segments. As summarized in the key, the symbolic representations are: black vertical lines of various thicknesses: locations of transposable elements, with thicker bars indicating clusters of elements; colored boxes: genes (both known and predicted) coded on one strand of DNA (boxes above the midline) and genes coded on the other strand (boxes below the midline). The length of each gene box includes both its exons (protein-coding DNA) and its introns (noncoding DNA) (see Figure 4–15); its height is proportional to the number of known cDNAs that match the gene. (As described in Chapter 8, cDNAs are DNA copies of mRNA molecules, and large collections of the nucleotide sequences of cDNAs have been deposited in a variety of databases, the more matches, the higher the confidence that the predicted gene is transcribed into RNA and is thus a genuine gene.) The color of each gene box indicates whether a closely related gene is known to occur in other organisms. For example, MWY means the gene has close relatives in mammals, in the nematode worm Caenorhabditis elegans, and in the yeast Saccharomyces cerevisiae. MW indicates the gene has close relatives in mammals and the worm but not in yeast. The rainbow-colored bar indicates percent G–C base pairs; across many different genomes, this percentage shows a striking regional variation, whose origin and significance are uncertain. (From M.D. Adams et al., Science 287:2185–2195, 2000. With permission from AAAS.)

The DNA in genomes does not direct protein synthesis itself, but instead uses RNA as an intermediary. When the cell needs a particular protein, the nucleotide sequence of the appropriate portion of the immensely long DNA molecule in a chromosome is first copied into RNA (a process called transcription). It is these RNA copies of segments of the DNA that are used directly as templates to direct the synthesis of the protein (a process called translation). The flow of genetic information in cells is therefore from DNA to RNA to protein (Figure 6–2). All cells, from bacteria to humans, express their genetic information in this way—a principle so fundamental that it is termed the central dogma of molecular biology.

Despite the universality of the central dogma, there are important variations in the way in which information flows from DNA to protein. Principal among these is that RNA transcripts in eucaryotic cells are subject to a series of processing steps in the nucleus, including RNA splicing, before they are permitted to exit from the nucleus and be translated into protein. These processing steps can critically change the "meaning" of an RNA molecule and are therefore crucial for understanding how eucaryotic cells read their genomes. Finally, although we focus on the production of the proteins encoded by the genome in this chapter, we see that for many genes RNA is the final product. Like proteins, many of these RNAs fold into precise three-dimensional structures that have structural, catalytic, and regulatory roles in the cell.

We begin this chapter with the first step in decoding a genome: the process of transcription by which an RNA molecule is produced from the DNA of a gene. We then follow the fate of this RNA molecule through the cell, finishing when a correctly folded protein molecule has been formed. At the end of the chapter, we consider how the present quite complex scheme of information storage, transcription, and translation might have arisen from simpler systems in the earliest stages of cell evolution.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6–3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most commonly by controlling the production of its RNA.

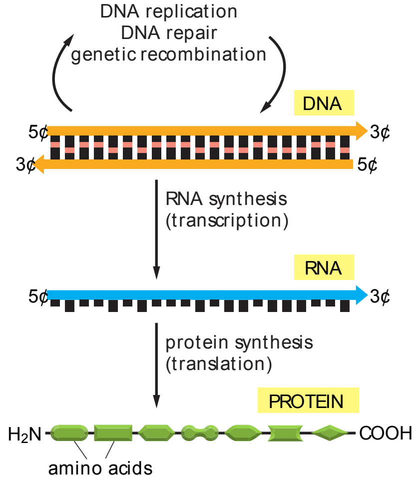

Figure 6–2 The pathway from DNA to protein. The flow of genetic information from DNA to RNA (transcription) and from RNA to protein (translation) occurs in all living cells.