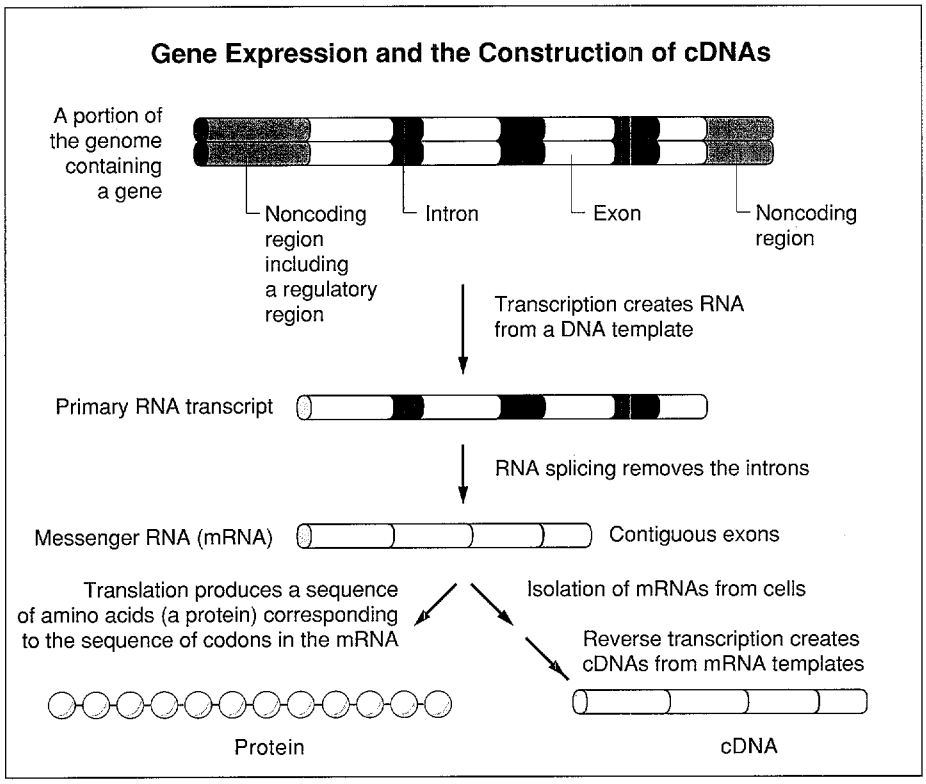
cDNAs and Expressed Genes

Copy DNAs, or cDNAs, are being synthesized, cloned, and sequenced as a source of STSs, unique landmarks for the physical map of the human genome. A cDNA is a copy of the protein-coding regions (exons) of a gene. It is not made directly from DNA isolated from the genome but rather, as shown in the figure, from the messenger RNA, the template that is translated into a protein. These templates are valuable because, unlike genomic DNA, each mRNA is a continuous stretch of protein-coding nucleotides. Moreover, the existence of an mRNA is proof that the corresponding protein-coding gene is an active, or expressed, gene.



cDNAs are synthesized in vitro. First, mRNAs are isolated from a population of tissue-specific cells. The isolated mRNAs represent only those genes that are being expressed in those particular cells. Each mRNA serves as a template in the synthesis of a complementary strand of DNA—the cDNA. The process of transcribing RNA into DNA, known as reverse transcription, is catalyzed by reverse transcriptase, an enzyme isolated from retroviruses, namely, RNA tumor viruses. The synthesized cDNAs are often shorter than the mRNA templates because of various processes that either degrade the mRNA or result in incomplete transcription. (Note that reverse transcriptase is not made by human cells. However, retroviruses, such as HIV, carry reverse transcriptase with them when they enter a host cell. The enzyme converts the viral RNA genome to DNA, which is then permanently incorporated into the genome of the host cell.)

After being synthesized in vitro, cDNAs are cloned. Cloned cDNAs have long been used for two purposes. First, cDNA libraries (random collections of cloned cDNAs) are used as sources of probes to identify the location of protein-coding regions in cloned fragments of genomic DNA. Second, particular mRNAs are isolated, converted to cDNAs, cloned, and then sequenced to determine the amino-acid sequence of the protein specified by the corresponding protein-coding gene.

The new emphasis is on sequencing short sections of cDNAs. If such a sequence is unique, it can be developed into a special kind of STS—one that is not only a unique, detectable landmark on the physical map of the genome but is also known to lie within an expressed gene. Furthermore, the cDNA sequence data provides some information about the protein encoded by the corresponding gene. ■