acids. The role of introns is not well understood, but they likely play a role in either evolution or cellular regulation or both. The presence of introns complicates the transfer of eucaryotic genes to protein production systems in procaryotes such as *E. coli*.

Two other *m*-RNA processing steps occur in eucaryotic cells that do not occur in procaryotes. One is *RNA capping,* in which the 5′ end is modified by the addition of a guanine nucleotide with a methyl group attached. The other is *polyadenylation,* in which a string of adenine nucleotides are added to the 3′ end. This tail of adenines is often several hundred nucleotides long. These two modifications are thought to increase *m*-RNA stability and facilitiate transport across the nuclear membrane.

Once we understand transcription, we can tackle translation. The translation of information on *m*-RNA into proteins occupies a very large fraction of the cells' resources. Like a large automobile plant, the generation of blueprints and the construction of the manufacturing machinery is worthless until product is made.

4.5. TRANSLATION: MESSAGE TO PRODUCT

4.5.1. Genetic Code: Universal Message

The blueprint for any living cell is the genetic code. The code is made up of three-letter words (*codons*) with an alphabet of four letters. Sixty-four words are possible, but many of these words are redundant. Although such a "language" may appear to be ridiculously simple, it is sufficient to serve as a complete blueprint for the "construction" of the reader.

The dictionary for this language is given in Table 4.1, and an illustration of the relationship of nucleotides in the chromosome and *m*-RNA to the final protein product is given in Fig. 4.6. The code is *degenerate* in that more than one codon can specify a particular amino acid (for example, UCU, UCC, UCA, and UCG all specify serine). Three codings, UAA, UAG, and UGA, are called *nonsense codons* in that they do not encode normally for amino acids. These codons act as stop points in translation and are encoded at the end of each gene.

The genetic code is essentially universal, although some exceptions exist (particularly in the mitochondria and for inclusion of rare amino acids). This essential universality greatly facilitates genetic engineering. The language used to make a human protein is understood in *E. coli* and yeast, and these simple cells will faithfully produce the same amino acid sequence as a human cell.

Knowing the genetic language, we may ask more about the mechanism by which proteins are actually made.

4.5.2. Translation: How the Machinery Works

The process of translation consists of three primary steps: *initiation*, *elongation*, and *termination*.

For initiation, *m*-RNA must bind to the ribosome. All protein synthesis begins with a AUG codon (or GUG) on the *m*-RNA. This AUG encodes for a modified methionine, *N*-formylmethionine. In the middle of a protein, AUG encodes for methionine, so the question is how the cell knows that a particular AUG is an initiation codon for