on their molecular types that include antibody-based drugs, Fc fusion proteins, anticoagulants, blood factors, bone morphogenetic proteins, engineered protein scaffolds, enzymes, growth factors, hormones, interferons, interleukins, and thrombolytics. They can also be classified based on their molecular mechanism of activity as (a) binding noncovalently to target, for example, mAbs; (b) affecting covalent bonds, for example, enzymes; and (c) exerting activity without specific interactions, for example, serum albumin. Most protein therapeutics currently on the market are recombinant, and hundreds of them are in clinical trials for the therapy of cancers, immune disorders, infections, and other diseases. New engineered proteins, including bispecific mAbs and multispecific fusion proteins, mAbs conjugated with small-molecule drugs, and proteins with optimized pharmacokinetics, are currently under development. Despite the remarkable growth in this category of drugs, the technology for their production remains genetic-engineering-based recombinant production. Perhaps novel techniques of the future may make it possible to synthesize these drugs, which may reduce some complexity, but that seems far; the next generation of biosimilars, as reported in Chapter 2, will likely be recombinant proteins expressed in prokaryotic and eukaryotic systems, the living systems that inevitably and invariably introduce significant variability in the primary, secondary, tertiary, and quaternary structures of these proteins. A keen understanding of the possible differences and their source is essential to develop biosimilars; this chapter provides this discussion.

9.2 Protein Structure

9.2.1 Building Elements

The 20 different naturally occurring amino acids give a staggering number of different possible proteins, 20^n to be exact, where n is the number of amino acid units or *residues* (Figures 9.2 and 9.3).

Each amino acid has a carboxylic group and an amine group, and amino acids link to one another to form a chain by a dehydration reaction by joining the carboxyl group of one amino acid with the amino group of the next. Thus, polypeptide chains have an end with an unbound carboxyl group, the *C*-terminus, and a beginning with an amine group, the *N*-terminus.

9.2.2 Translation

When a protein is translated from nRNA, it is created from the *N*-terminus to the *C*-terminus. The amino end of an amino acid (on a charged transfer RNA [tRNA]) during the elongation stage of translation attaches to the carboxyl end of the growing chain. Since the start codon of the genetic code codes for the amino acid methionine, most protein sequences start with a methionine (or, in bacteria, mitochondria, and chloroplasts, the modified version *N*-formylmethionine [fMet]). However, some proteins are modified posttranslationally, for example, by cleavage from a protein precursor, and, therefore, they may have different amino acids at their *N*-terminus.