14.9. PROTEIN ENGINEERING

Not only can cells be engineered to make high levels of naturally occurring proteins or to introduce new pathways, but we can also make novel proteins. It is possible to make synthetic genes encoding for totally new proteins. We are beginning to understand the rules by which a protein's primary structure is converted into its three-dimensional form. We are just learning how to relate a protein's shape to its functional properties, stability, and catalytic activity. In the future, it may become possible to customize protein design to a particular well-defined purpose.

Protein engineering at present mainly involves the modification of existing proteins to improve their stability, substrate and inhibitor affinity and specificity, and catalytic rate. Generally, the protein structure must be known from x-ray crystallography. Key amino acids in the structure are selected for alteration based on computer modeling, on interactions of the protein with substrates, or by analogy to proteins of related structure. The technique used to generate genes encoding the desired changes in protein structure is called *site-directed mutagenesis* (see Fig. 14.9). Using this approach, any desired amino acid can be inserted precisely into the desired position.

The reader may wonder why site-directed mutagenesis would be preferred to simple mutation-selection procedures. One reason may be that mutation followed by selection for particular properties may be difficult when the alterations in protein properties are subtle and confer no advantages or disadvantages on the mutant cell. A second reason is that site-directed mutagenesis can be used to generate the insertion of an amino acid in a particular location, while a random mutation giving the same result would occur so infrequently as to be unobtainable.

To make this point more evident, consider the degenerate nature of the genetic code. Each codon consists of three letters. The odds for mutation in one of these three letters is about 10^{-8} per generation. The odds that two letters would simultaneously be altered is much lower (order of 10^{-16}). The codon UAC (for tyrosine) could be altered by single-letter substitutions to give AAC (asparagine), GAC (aspartate), CAC (histidine), UCC (serine), UUC (phenylalanine), UGC (cysteine), UAA (stop signal), UAG (stop signal), and UAU (tyrosine). Random mutants in this case are very unlikely to carry substitutions for 13 of the 20 amino acids. Thus, most of the potential insertions can be generated reliably only by using site-directed mutagenesis.

The above approaches are directed toward the rational design of proteins. An alternative, and often complementary, approach is that of *directed evolution.* This process is based on random mutagenesis of a gene and the subsequent selection of proteins with desired properties. Large libraries of mutant genes must be made so that the rare beneficial forms are present. A rapid screen or selection must be available to select those mutants with the desired function or characteristics. One technique to generate mutants is the use of "error-prone" PCR. Another approach is DNA shuffling, which requires genes from homologous proteins. Segments from the genes are recombined randomly to form chimeric genes. Proteins are typically selected for improved stability, binding strength, catalytic activity, or solubility. In some cases, the screen is for new activities based on ability to bind other molecules not normally bound by the native protein. These techniques combined with improved biochemical methods are leading to a better understanding of the relation of protein structure to function. Also, these techniques support the extension of protein