

Protein engineering

The design and construction of new proteins or enzymes with novel or desired functions, through the modification of amino acid sequences using recombinant DNA (deoxyribonucleic acid) technology. The sizes and three-dimensional conformations of protein molecules are also manipulated by protein engineering. The basic techniques of genetic engineering are used to alter the genes that encode proteins, generating proteins with novel activities or properties. Such manipulations are frequently used to discover structure-function relationships, as well as to alter the activity, stability, localization, and structure of proteins. *See* GENETIC ENGINEERING; PROTEIN.

Point mutants. Many subtle variations in a particular protein can be generated by making amino acid replacements at specific positions in the polypeptide sequence. Each protein is unique by virtue of the sequence of its amino acids. At any position in the sequence, an amino acid can be replaced by another to generate a mutant protein that may have different characteristics by virtue of the single replaced amino acid. For example, pancreatic ribonuclease A is an enzyme comprising 124 amino acids that cleaves the covalent bonds that join ribonucleic acids (RNA). If at position 119 in the sequence the naturally occurring histidine is replaced with an alanine, the mutant protein is referred to as a histidine 119 → alanine (H119A) mutant of ribonuclease A. This mutant protein is expected to have little or no biological activity, because histidine 119 is important for that activity. Other mutations have very little effect on their proteins. This is particularly true when the amino acid being changed is substituted with other closely related amino acids and when the amino acid is not conserved in the same protein found in other organisms. Typically, site-directed protein engineering targets amino acids that are involved in a particular biological activity. *See* AMINO ACID; MUTATION.

Because there are 20 different amino acids, 20 different variants of ribonuclease A can be created just by having a set that differs only by the amino acid at position 119. If changes are also made at another position, for example, position 41 (where normally there is a lysine), then in principle there are 400 different variants that can be created by making all possible combinations of substitutions at just these two positions in the sequence. As more positions are varied, the number of combinations becomes enormous so that, with just six different positions subjected to all possible variations, 64 million different proteins can be generated. These proteins will share the identical sequence, except at the six varied positions.

Deletion mutants. In addition to the substitution of amino acids at specific positions, amino acids can be deleted from the sequence, either individually or in groups. These proteins are referred to as deletion mutants. Deletion mutants may or may not be missing one or more functions or properties of the full, naturally occurring protein. Proteins from eukaryotes, such as mammals, plants, and fungi, can have very

long amino acid sequences and as a consequence tend to be organized as several modular protein domains linked together. Deletion mutants have been a useful way to create smaller proteins that contain only one or a few of these domains so that the individual properties can be studied.

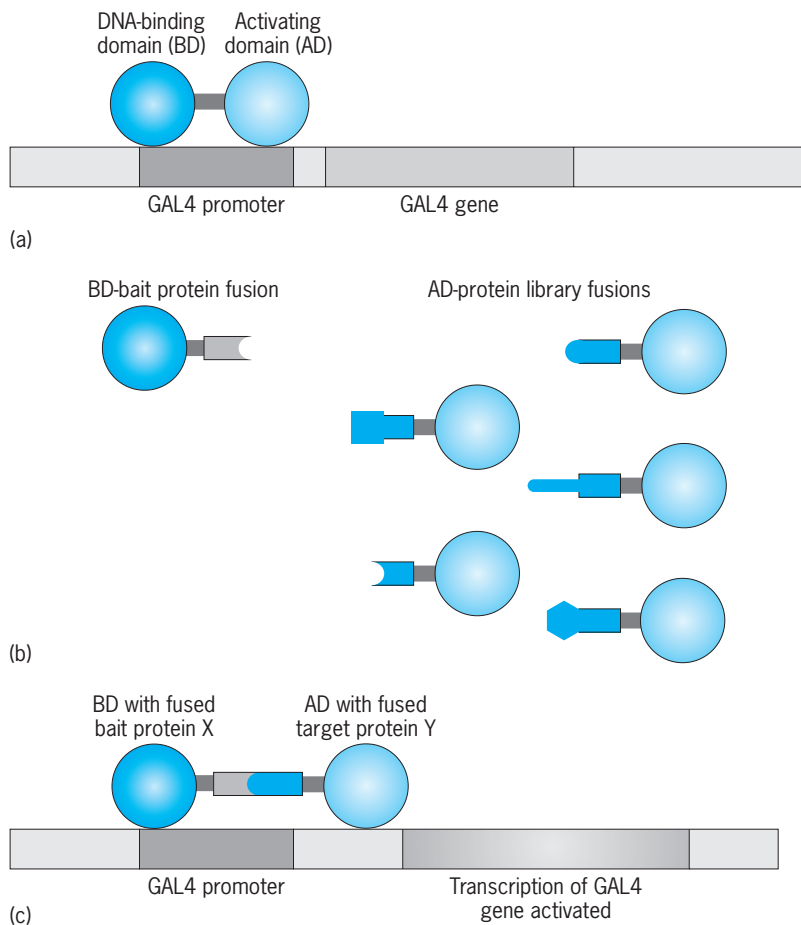
Hybrid/fusion proteins. Protein sequences can be joined or fused to that of another protein. The resulting protein is called a hybrid, fusion, or chimeric protein, which generally has characteristics that combine those of each of the joined partners. Protein fusions have been extensively used to study interactions between two or more proteins. For example, an application of protein fusion methods called the yeast two-hybrid screen was developed to identify proteins that interact with each other (*see illustration*). The system was developed by separating a yeast transcription-activating protein into two functional domains. The first domain of the transcription-activating protein is fused to a protein that is being studied for interactions. The second domain of the transcription-activating protein is fused to a collection of proteins encoded in a protein library. The first fusion protein and one from the protein library are expressed together in yeast. When the two fusion proteins interact, they can bring together the two halves of the transcription activator and turn on the expression of a reporter gene (*see illustration*). Another example is the fusion of a protein to an inherently fluorescent protein, such as the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*. Fluorescent fusion proteins have been used to study the location, movement, appearance, and degradation of proteins in living cells. *See* FLUORESCENCE.

Another common application for fusion proteins is to facilitate purification by affinity chromatography. A short sequence of amino acids or a protein domain is fused to one end of the polypeptide. Six histidines can be used to purify a protein by affinity to nickel. The Fc region of an antibody (region of an antibody molecule that binds to cell-surface antibody receptors) can be fused to a protein to purify it by affinity to protein A, a protein from bacteria that specifically binds antibodies. *See* LIQUID CHROMATOGRAPHY.

Novel activities or properties. Point and deletion mutants and hybrid proteins are constructed to obtain polypeptides with new properties. These proteins are either created individually, by site-directed mutagenesis, or they are generated as a large pool or library of millions of variants. The library is then screened or subjected to a special selection procedure to obtain the protein or proteins with the desired characteristics.

Proteins generally have stable and unique three-dimensional structures and are active at physiological temperatures of 37°C (98.6°F). They usually lose their three-dimensional shapes as the temperature is raised more than 5 or 10° above 37°C and, as a consequence, lose their biological activities. In some instances, a protein may be cold-sensitive; that is, it may lose its conformation and activity at lower

2 Protein engineering



Yeast-two-hybrid screen. (a) The yeast transcription-activating protein GAL4 has two separate domains: a DNA-binding domain (BD), which binds the promoter region of the GAL4 gene, and a transcription-activating domain (AD), which stimulates production of the GAL4 RNA transcript. (b) These two domains are modular and can be fused onto other proteins while retaining their function. This forms the basis of the yeast two-hybrid assay (screen). The first hybrid, or fusion, protein consists of a protein of interest (bait protein) fused to the GAL4 BD. The second hybrid is a collection of fusion proteins made between the GAL4 AD and a library of many proteins. (c) If a protein from the library is able to interact with the first fusion protein, visualized in the figure as a protein with a complementary shape, the GAL4 transcription factor is reconstituted, and transcription of the gene is activated. Proteins that do not interact, those with noncomplementary shapes, fail to activate the system.

temperatures. The effect of temperature on protein stability can be modified by protein engineering, in particular, by introducing amino acid replacements that enhance or destabilize the molecular packing interactions in the core of the protein structure, such as has been done with the enzymes T4 lysozyme and staphylococcal nuclease.

Protein engineering has been used to produce therapeutic proteins with improved properties such as increased solubility and stability. For example, insulin was engineered through mutagenesis to create monomeric forms that are fast acting (insulin lispro and insulin aspart). Conversely, another form (insulin glargine) was created by mutagenesis to precipitate upon injection and give a sustained release of insulin. Mutation of a free cysteine in aldesleukin (a synthetic version of interleukin-2 used to treat some forms of cancer) or interferons beta-1b (used to treat relapsing-remitting multiple sclerosis) was used to produce therapeutics with decreased aggrega-

tion. See CANCER (MEDICINE); INSULIN; INTERFERON; INTERLEUKIN; MULTIPLE SCLEROSIS.

Proteins can be engineered to acquire new biological activities. For example, a catalytic antibody is a variant of an antibody. Antibodies are proteins that normally bind to a specific molecule but do not alter the bound molecule in any way. A catalytic antibody is one which has been changed by mutations to have a novel sequence that folds into a structure that catalyzes a specific reaction (such as amide bond formation, ester hydrolysis, and decarboxylation). Catalytic antibodies function like enzymes, and are created to catalyze reactions for which there are no naturally occurring enzymes. Fifty or more different reactions have been catalyzed by the action of catalytic antibodies that were obtained individually by methods of protein engineering. See ANTIBODY; CATALYTIC ANTIBODY.

Structure-function relationships. With over 180 genomes sequenced, including human and mouse, the amino acid sequence of a particular protein is now generally available from many different organisms. This advance during the past decade has created a situation in which protein sequences are available even when there is no information on the biological activity of the protein. Thus, the focus has shifted from sequencing to understanding the function of all of these proteins. Techniques such as the yeast two-hybrid system have been applied to identify protein interactions, and crystallographic structures are being determined to elucidate biological information about all the proteins in an entire organism. Still, the chemical basis for activity is often not completely understood, even after the biological activity and high-resolution structure have been determined. The construction of mutant proteins can elucidate the role of a particular amino acid at a specific position in the sequence. See GENE.

Protein engineering has been applied to understand how enzymes catalyze reactions, from identifying which amino acids are essential for catalysis to analyzing how amino acid changes alter particular aspects of a reaction such as substrate specificity. Antibiotics such as erythromycin are made by large multidomain proteins called polyketide synthases. Site-directed mutagenesis has been used to modify the substrate specificity of one polyketide synthase reaction so that the product contains a malonate unit, whereas the product of the original enzyme contained a methylmalonate unit. In addition to site-directed mutagenesis, the order of the polyketide synthase domains have been shuffled to create proteins that could catalyze the synthesis of new antibiotics. See ANTIBIOTIC; ENZYME.

An extension of site-directed mutagenesis allows nonnatural amino acids to be incorporated into proteins. Nonnatural amino acids are not naturally encoded by the genome, but instead include a wide variety of amino acids that are present in cells or produced by synthetic methods. This protein engineering method allows the chemical properties of a particular amino acid to be changed beyond that naturally encoded by a gene. For example, nonnatural

amino acids incorporated into proteins may contain sugars, nucleophiles (electron donors), electrophiles (electron acceptors), crosslinkers (a chemical compound that forms covalent bonds between adjacent polymer chains that lock the chains in place), or altered shapes and sizes. Detailed studies on ion channels with over 60 nonnatural amino acids incorporated illustrate the chemical power of this approach.

Minimalist proteins. Proteins are usually large molecules composed of several hundred or even more than a thousand amino acids. Yet, the portion of the protein responsible for a specific biological activity is usually concentrated in a small part of the structure. In these circumstances, there are advantages to creating a minimalist protein having enough of the structure to retain the desired activity. A smaller protein is often superior for analyzing and manipulating structure-function relationships, and the reduced cost of materials for manufacturing a small protein can also be an important motivation for reducing the size.

To obtain minimalist proteins with novel properties, libraries with large numbers of different amino acid sequences have been created. These sequences

are often so short (less than 50 amino acids) that they are referred to as peptides, oligopeptides, or polypeptides, and not as proteins. (A protein is regarded as a long polypeptide, generally at least over 80 amino acids in length and usually much longer.) Those with novel properties, such as an ability to bind to a specific ligand (a unique molecule that is usually small) are then selected. In these instances, the sequences of the peptides are created de novo and are not based on that of a naturally occurring protein. *See* GENE; GENE ACTION; GENETIC CODE; PEPTIDE.

Karla L. Ewalt; Paul Schimmel

Bibliography. J. A. Branningan and A. J. Wilkinson, Protein engineering 20 years on, *Nature Rev. Mol. Cell Biol.*, 3:964-970, 2002; J. Lippincott-Schwartz and G. H. Patterson, Development and use of fluorescent protein markers in living cells, *Science*, 300:87-91, 2003; S. A. Marshall et al., Rational design and engineering of therapeutic proteins, *Drug Discovery Today*, 8:212-221, 2003; P. G. Schultz et al., The chemistry of the antibody molecule, *Angew. Chem. Int. Ed. Engl.*, 41:4427-4437, 2002; L. Wang and P. G. Schultz, Expanding the genetic code, *Chem. Commun.*, 1:1-11, 2002.