Kirk-Othmer Encyclopedia of Chemical Technology.

PROTEIN ENGINEERING

Protein engineering encompasses a wide variety of techniques, ranging from the rational modification of existing proteins to the de novo design of novel proteins. Protein engineering is most commonly carried out by manipulation of the protein at the genetic level; that is, by mutagenesis of the gene (deoxyribonucleic acid (DNA)) which codes for the protein (see Genetic engineering; Nucleic acids). Other methods, such as chemical modification and chemical synthesis of proteins (qv), are also included within the scope of protein engineering.

One of the principal goals of protein engineering is to provide insight into the complex relationship between protein structure and function. Many protein engineering studies have involved changing specific amino acid residues within proteins of known crystal structures in order to test predictions as to how such changes may affect the structure or function of the protein. Even in the absence of a well-defined three-dimensional structure, mutagenesis studies can provide a wealth of information on the importance of a particular amino acid residue to various physicochemical properties of the protein, such as stability, enzyme catalysis, ligand binding, and protein–protein interactions.

Protein engineering encompasses a vast amount and wide variety of research. At least two textbooks (1, 2) have been devoted exclusively to this topic, and several excellent reviews have been published (3, 4). Herein, an overview of principles, an introduction to basic techniques, and a summary of results of representative experiments on protein engineering are provided.

1. Protein Structure

Much of protein engineering concerns attempts to explore the relationship between protein structure and function. Proteins are polymers of amino acids (qv), which have general structure $+H_3N—CHR—COO^-$, where R, the amino acid side chain, determines the unique identity and hence the structure and reactivity of the amino acid (Fig. 1, Table 1). Formation of a polypeptide or protein from the constituent amino acids involves the condensation of the amino-nitrogen of one residue to the carboxylate-carbon of another residue to form an amide, also called peptide, bond and water. The linear order in which amino acids are linked in the protein is called the primary structure of the protein or, more commonly, the amino acid sequence. Only 20 amino acid structures are used commonly in the cellular biosynthesis of proteins (qv).

Discrete segments of the polypeptide chain can fold into regular, repeating structural motifs, known as secondary structures. It is the identity and sequence of amino acids within a polypeptide chain that dictate its secondary structure. As of 1996, however, understanding of the factors involved in protein folding is not sufficient to predict accurately the pattern of secondary structure from the amino acid sequence of a protein. The two most common structural motifs in proteins, the α-helix and β-pleated sheet, are shown in Figure 2. Short segments of sequence that allow the contiguous polypeptide to change directions (β-turns), comprise another important secondary structure element in proteins. Regions of the protein that do not contain regular, ie, ordered, secondary structure are referred to as random coil.

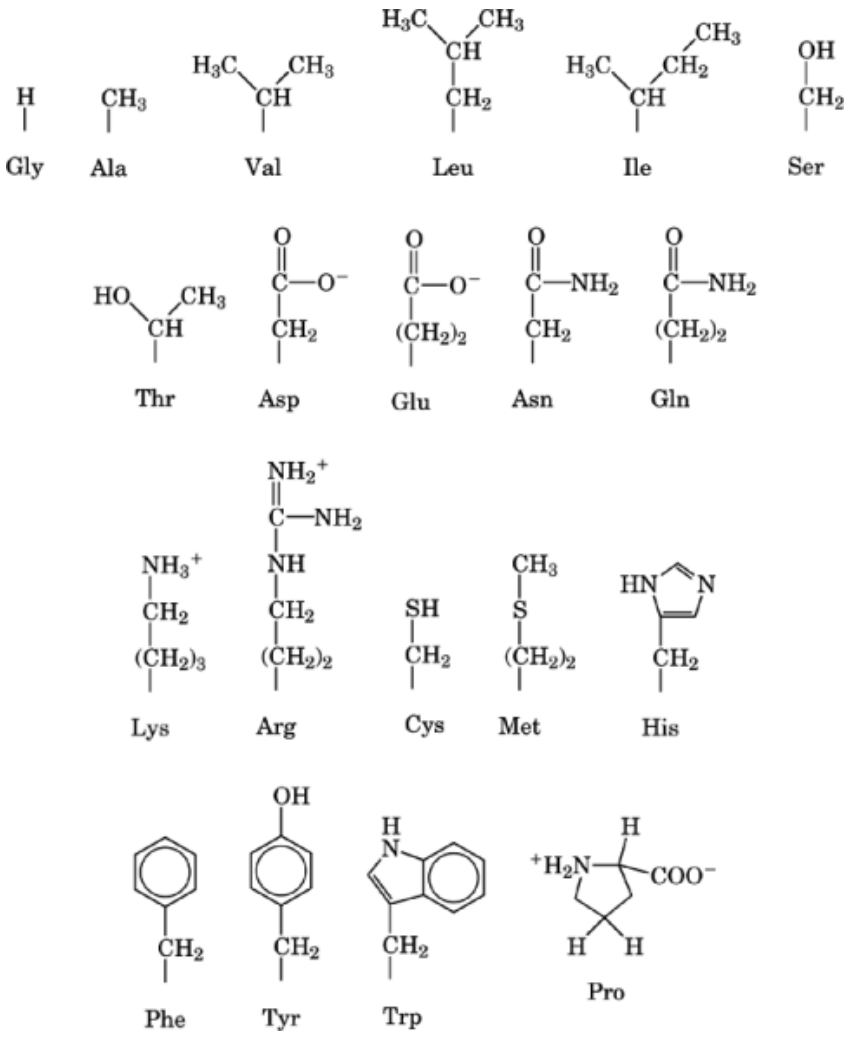

Fig. 1. The side chain R of the 20 standard amino acids $^{+}H_3N–CHR–COO^{-}$ at pH 7. For proline, the complete structure is shown. Amino acid side chains can be categorized as aliphatic (Gly, Ala, Val, Leu, and Ile), hydrophilic (Ser, Thr, Asp, Glu, Asn, Gln, Lys, and Arg), sulfur-containing (Cys and Met), aromatic (His, Phe, Tyr, and Trp), and the imino acid proline.

The way in which the elements of secondary structure fold upon each other to form compact globular structures is referred to as tertiary structure. Noncovalent forces such as hydrogen bonds, electrostatic interactions, and hydrophobic interactions play a dominant role in determining protein tertiary structure. In some proteins, covalent bonds formed by the oxidation of two cysteine sulfhydryl groups provide additional stability to the folded protein. These cross-links are referred to as disulfide bonds. It is the tertiary structure that gives a protein its overall shape and dimensions. Tertiary structure also provides a means of bringing into spatial proximity amino acids that may be distant in the linear sequence. Often the active form of a protein is a complex of polypeptides held together by noncovalent or covalent (disulfide) interactions. The arrangement of these polypeptide subunits relative to one another defines the quaternary structure of the protein. These

Table 1. Physicochemical Properties of the Natural Amino Acids[a]

Amino acid	Three-letter code	One-letter code	Mass of residue in proteins[b]	Accessible surface area[c], nm^2	Hydrophobicity index[d]	pK_a of ionizable side chain	Occurrence in proteins, %[e]	Relative mutability[f]
alanine	Ala	A	71.08	1.15	+1.8		9.0	100
arginine	Arg	R	156.20	2.25	−4.5	12.5	4.7	65
asparagine	Asn	N	114.11	1.60	−3.5		4.4	134
aspartate	Asp	D	115.09	1.50	−3.5	3.9	5.5	106
cysteine	Cys	C	103.14	1.35	+2.5	8.4	2.8	20
glutamate	Glu	E	128.14	1.80	−3.5	4.1	3.9	102
glutamine	Gln	Q	129.12	1.90	−3.5		6.2	93
glycine	Gly	G	57.06	0.75	−0.4		7.5	49
histidine	His	H	137.15	1.95	−3.2	6.0	2.1	66
isoleucine	Ile	I	113.17	1.75	+4.5		4.6	96
leucine	Leu	L	113.17	1.70	+3.8		7.5	40
lysine	Lys	K	128.18	2.00	−3.9	10.8	7.0	56
methionine	Met	M	131.21	1.85	+1.9		1.7	94
phenylalanine	Phe	F	147.18	2.10	+2.8		3.5	41
proline	Pro	P	97.12	1.45	−1.6		4.6	56
serine	Ser	S	87.08	1.15	−0.8		7.1	120
threonine	Thr	T	101.11	1.40	−0.7		6.0	97
tryptophan	Trp	W	186.21	2.55	−0.9		1.1	18
tyrosine	Tyr	Y	163.18	2.30	−1.3	10.1	3.5	41
valine	Val	V	99.14	1.55	+4.2		6.9	74

[a](Courtesy of VCH Publishers, Inc. (5).)
[b]Values reflect the molecular weights of the amino acids minus that of water.
[c]For residues as part of a polypeptide chain (6).
[d]Ref. 7.
[e]Based on the frequency of occurrence for each residue in the sequence of 207 unrelated proteins (8).
[f]Relative mutability represents the likelihood that a residue will mutate within a specified time period during evolution. Units are arbitrary, with alanine assigned a value of 100 (9).

different levels of protein structure, ie, primary, secondary, tertiary, and quaternary structure, are summarized in Figure 3.

1.1. Noncovalent Forces Stabilizing Protein Structure

Much of protein engineering concerns attempts to alter the structure or function of a protein in a predefined way. An understanding of the underlying physicochemical forces that participate in protein folding and structural stabilization is thus important.

Through combined effects of noncovalent forces, proteins fold into secondary structures, and hence a tertiary structure that defines the native state or conformation of a protein. The native state is then that three-dimensional arrangement of the polypeptide chain and amino acid side chains that best facilitates the biological activity of a protein, at the same time providing structural stability. Through protein engineering subtle adjustments in the structure of the protein can be made that can dramatically alter its function or stability.

1.1.1. Electrostatics

Electrostatic interactions, such as salt bridges, result from the electrostatic attraction that occurs between oppositely charged molecules. These usually involve a single cation, eg, the side chain of Lys or Arg, or the

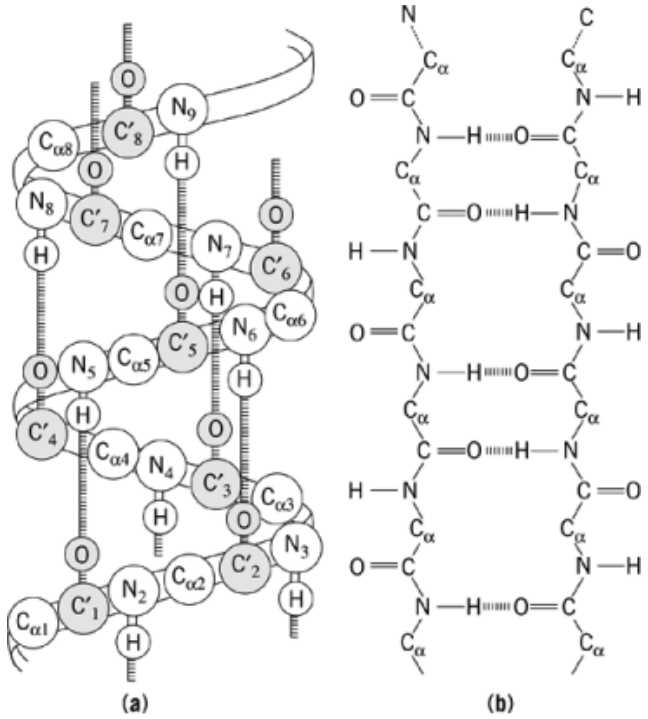

Fig. 2. Protein secondary structure: (**a**) the right-handed α-helix, stabilized by intrasegmental hydrogen-bonding between the backbone CO of residue i and the NH of residue i+4 along the polypeptide chain. Each turn of the helix requires 3.6 residues. Translation along the helical axis is 0.15 nm per residue, or 0.54 nm per turn; and (**b**) the β-pleated sheet where the polypeptide is in an extended conformation and backbone hydrogen-bonding occurs between residues on adjacent strands. Here, the backbone CO and NH atoms are in the plane of the page and the amino acid side chains extend from C_α; alternating above and below the plane of the page (10).

amino terminus, etc, interacting with a single anion, eg, the side chain of Glu or Asp, or the carboxyl terminus, etc. This attractive force is inversely proportional to the distance between the charges and the dielectric constant of the solvent, as described by Coulomb's law.

1.1.2. Hydrogen Bonds

An attractive force which involves the sharing of a hydrogen atom between two electronegative atoms, the hydrogen-bond (H-bond), consists of a hydrogen donor to which the H-atom is covalently bound and an acceptor containing either a partial or full negative charge that attracts the electropositive H-atom. Eleven of the 20 amino acids can participate in H-bonds, Arg, Asp, Asn, Cys, Glu, Gln, His, Lys, Ser, Thr, and Tyr. Stronger than van der Waals bonds but much weaker than covalent bonds, hydrogen bonds constitute a primary determinant of biological specificity. These bonds are sufficiently strong to direct molecular interactions such as the attraction between an enzyme and its substrate, but sufficiently weak to be reversibly made and broken, as in enzyme catalysis (11).

1.1.3. Van der Waals Interactions

Van der Waals interactions result from the asymmetric distribution of electronic charge surrounding an atom, which induces a complementary dipole in a neighboring atom, resulting in an attractive force. In general,

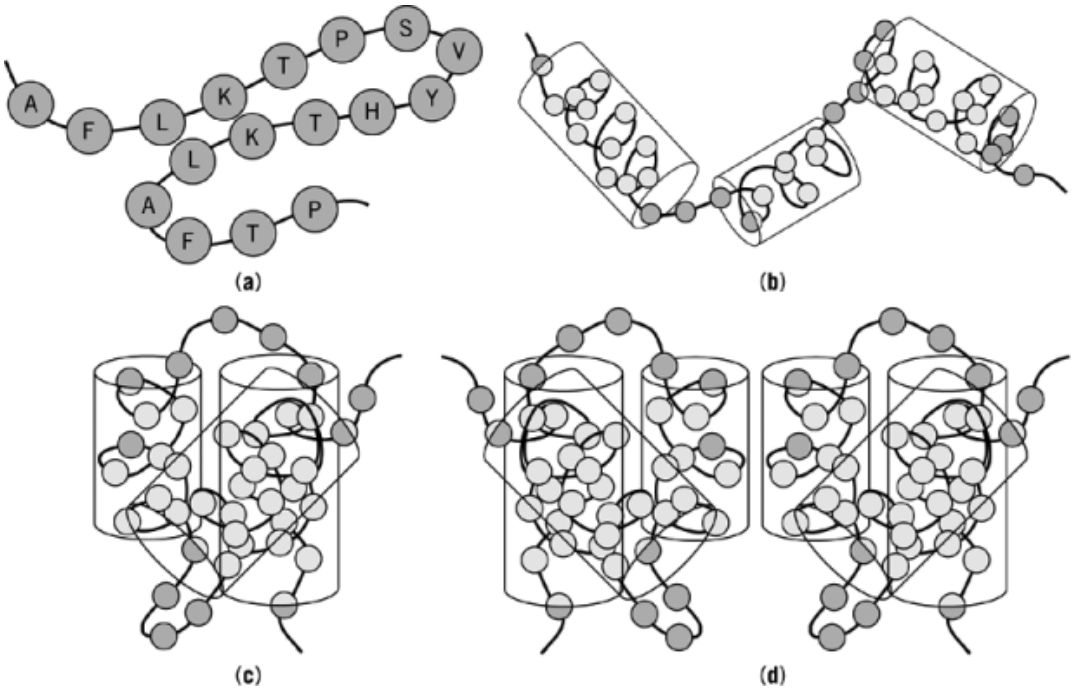

Fig. 3. The hierarchy of protein structures: (**a**) primary structure (see Table 1 for amino acid code); (**b**) secondary structure; (**c**) tertiary structure; and (**d**) quaternary structure (10).

the attractive force of van der Waals interactions is very weak ($<$4.2 kJ/mol (1 kcal/mol)) but may become significant if steric complementarity creates an opportunity to form a large number of van der Waals attractions.

1.1.4. The Hydrophobic Effect

Noncovalent interactions can occur not only between different parts of the polypeptide chain but also between the polypeptide and the solvent water. Water very effectively solvates polar molecules, weakening electrostatic forces and H-bonds by competing for the elements of those attractions. Nonpolar amino acids cannot, however, participate in such favorable interactions with water. This absence of interaction, coupled with the high affinity of water for itself, forces the nonpolar molecules to associate, forming a sequestered, hydrophobic core in the interior of the protein. This process is termed the hydrophobic effect. It is generally accepted that hydrophobic attractions are the principal driving force in protein folding, primarily owing to the much more favorable entropy of the water resulting from exclusion of nonpolar molecules.

1.2. Techniques Used to Study Protein Structure

The structural features of proteins can be experimentally assessed by a variety of biophysical methods. Spectroscopic methods can assess aspects of secondary and tertiary structure (see Spectroscopy, optical) (10); chromatographic and hydrodynamic methods can provide information on subunit assembly (see Chromatography). Protein structure at the atomic level can be assessed by x-ray crystallography (12) (see X-ray technology) and by multidimensional nmr spectroscopy (see Magnetic spin resonance) (13). These powerful structural methods have played critical roles in protein engineering. Discussions of these techniques may also be found in various journals, such as *Nature*, *Structural Biology* (*London*), and *Current Opinion in Structural Biology*.

2. The Protein Engineering Process

2.1. Identification and Characterization of the Target Protein

Prior to commencement of the process of protein engineering, a target protein must be selected and some structural information on the natural form of that protein determined. The motivation for selecting a particular target protein varies according to the researcher and the problem to be solved. Once the target protein has been selected, a concerted effort must be made to obtain the natural protein in pure form in order to determine the needed structural information. At a minimum, some amino acid sequence information must be known in order to design the oligonucleotide probes that are used to identify the gene for the target protein. Additionally, information on other properties of the target protein, such as molecular weight and isoelectric point are useful to confirming the identity of the cloned protein at a later date. Knowing what, if any, post-translational modifications (14) occur for the natural protein can guide the choice of an appropriate host organism for expression, and thus it is important to obtain this information prior to attempting cloning and expression.

Whereas most proteins are soluble in the cell cytosol, many proteins are associated with the various membranes found in cells. Membrane proteins present additional challenges and one must therefore know in advance whether or not the target protein is membrane-associated. Finally, it is important to know whether the active form of the target protein is a single, monomeric polypeptide, or if it consists of multiple subunits. Which it is greatly affects the options available for cloning and expression strategies. Methods of protein analysis have been established that can help to answer these and other questions about the nature of the target protein. Descriptions of these various methods can be found in any of a number of texts devoted to protein analysis (10, 15).

2.2. Cellular Protein Biosynthesis

The process of cellular protein biosynthesis is virtually the same in all organisms. The information which defines the amino acid sequence of a protein is encoded by its corresponding sequence of DNA (the gene). The DNA is composed of two strands of polynucleotides, each comprising some arrangement (sequence) of the four nucleotide building blocks of the nucleic acids: adenine (A), thymine (T), guanine (G), and cytosine (C). In ribonucleic acid (RNA), T is replaced by uracil (U). The two strands of DNA are held together by hydrogen-bonding patterns between specific nucleotide bases, ie, base pairing. The structures of the bases are such that A always pairs with T, G with C. Hence, the two strands of DNA are complementary; that is, the nucleotide sequence of one strand defines exactly the sequence of the other strand. For example, if one strand had the sequence ATCGA, the complementary strand would have to have the sequence TAGCT. This complementarity provides a convenient means for living organisms to replicate their genetic information and forms the technical basis for protein engineering.

The relationship between the nucleotide sequence of the DNA and the amino acid sequence of the protein is known as the genetic code (Table 2). In this code, a sequence of three nucleotides (a codon) specifies a particular amino acid. The genetic code consists of 64 codons, of which 61 code for amino acids and three are translation stop signals (stop codons). Because the genetic code is nearly the same in all organisms, it is possible to insert genetic information from one organism into a different host organism. The latter can then use its biosynthetic machinery to produce the foreign protein. This process is referred to as heterologous protein expression.

Cellular protein biosynthesis involves the following steps. One strand of double-stranded DNA serves as a template strand for the synthesis of a complementary single-stranded messenger ribonucleic acid (mRNA) in a process called transcription. This mRNA in turn serves as a template to direct the synthesis of the protein in a process called translation. The codons of the mRNA are read sequentially by transfer RNA (tRNA) molecules, which bind specifically to the mRNA via triplets of nucleotides that are complementary to the particular codon,

Table 2. The Genetic Code

First position (5′-end)	Second position				Third position (3′-end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met (START)	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

called an anticodon. Protein synthesis occurs on a ribosome, a complex consisting of more than 50 different proteins and several structural RNA molecules, which moves along the mRNA and mediates the binding of the tRNA molecules and the formation of the nascent peptide chain. The tRNA molecule carries an activated form of the specific amino acid to the ribosome where it is added to the end of the growing peptide chain. There is at least one tRNA for each amino acid.

2.3. Cloning of the Target Protein

Cloning procedures have been made possible by the availability of several key types of enzymes, including restriction endonucleases, enzymes that cleave double-stranded DNA at specific sites; DNA polymerases (especially thermostable forms), which replicate DNA templates; and DNA ligases, which covalently link (ligate) fragments of DNA. In general, fragments of DNA are ligated into vectors that can autonomously replicate their DNA in the host organism. The two most common vectors used in bacteria are plasmids and phage. Plasmids are naturally occurring circular DNA molecules that can act as accessory chromosomes. Phage, such as lambda phage, are viruses which can both stably integrate into the chromosome of the host (lysogenic pathway) or use the host machinery to produce more viral particles, which in turn lyse the host cell and destroy it (lytic pathway). The M13 phage are single-stranded circular molecules of DNA that do not kill the host when these are packaged and secreted. A double-stranded replicative form of this phage also exists and can be manipulated like plasmids. An excellent introduction to recombinant DNA technology is available (16). Detailed protocols on methodology can be found in References 17 and 18.

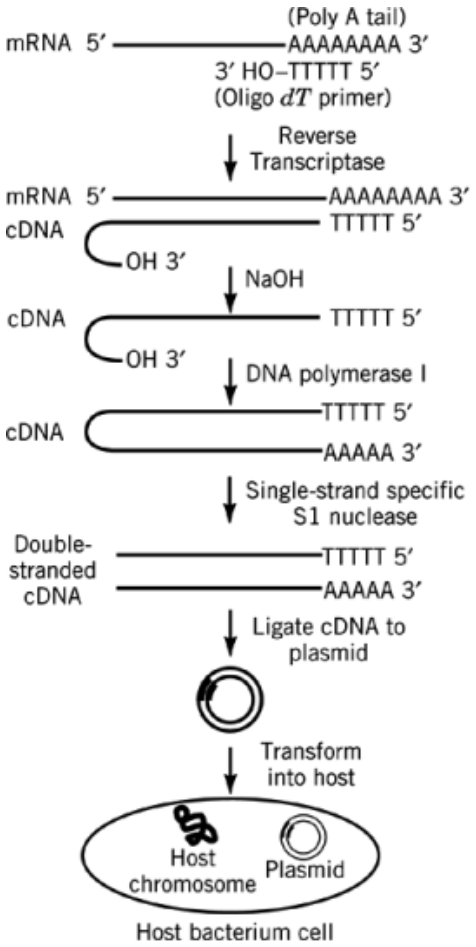

Fig. 4. Steps in making a cDNA library. Cellular mRNA is used as a template to make a complementary DNA. This cDNA is then ligated to a plasmid, which can be transformed into an appropriate host for propagation. See text.

2.3.1. Making a cDNA Library

Within the genes of eukaryotes, extraneous stretches of nucleotide sequences that do not code for amino acid residues in the protein (introns) are often interspersed among the coding sequences (exons). When DNA is transcribed, these introns are excised from the mRNA to form a contiguous sequence of exons that translate into the contiguous amino acid sequence of the protein. Eliminating the introns from the DNA directly would provide a convenient tool for cloning purposes. Therefore, complementary DNA (cDNA) libraries are constructed from cellular mRNA (Fig. 4). This form of RNA has already been processed by the cell and therefore the DNA corresponds to the actual nucleotide sequence that codes for the protein without the interrupting introns.

Most eukaryotic genes contain a signal sequence after the stop codon that causes further processing of the mRNA by the enzyme poly-A polymerase, adding about 250 A residues to the 3′ end of the newly transcribed mRNA. The procedure for making cDNA from mRNA takes advantage of this unique marker of the 3′-end of the gene by using a poly-T oligonucleotide primer to base-pair to the poly-A tail. This primer is then extended by the enzyme reverse transcriptase which, recognizing the nucleotide sequence of the mRNA, makes a complementary strand of DNA to produce a DNA–RNA hybrid. The RNA part of this duplex is removed by

treatment with sodium hydroxide, which hydrolyzes the RNA but leaves the DNA intact. The DNA polymerase is then used to make a complementary copy of the single-stranded DNA. After removal of the single-stranded hairpin loop from the duplex DNA by treatment with S1 nuclease (see step 4 of Fig. 4), the linear cDNA can be ligated to an appropriate plasmid vector and transformed into bacteria, where it is propagated. Because the cDNA was generated from a mixture of mRNAs, the cDNA is also a mixture of molecules. The transformation of the recombinant cDNA plasmids into the bacteria is designed so that each bacterial cell receives a unique cDNA recombinant.

2.3.2. Screening of a cDNA Library

The bacteria containing individual cDNA recombinant plasmids are grown on an agar plate where individual colonies can be visualized, and these colonies are screened for the presence of the desired DNA, using a probe specific to the DNA of interest. The probe is generally an oligonucleotide where the sequence complements a segment of the nucleotide sequence of the cDNA of interest. The probe is usually radiolabeled to allow easy identification of its location. The bacterial cells on an agar plate are replica-plated onto nitrocellulose filters, where the cells are immobilized. This process involves touching the nitrocellulose to the colonies on the agar plate so that some cells from each colony are transferred to the filter. The filter thus becomes an exact replica of the arrangement of colonies on the agar plate. The cells are lysed and the DNA inside is denatured into single-stranded forms. The probe, which is present in excess, can then hybridize (base-pair) to complementary DNA on the nitrocellulose. The location of the hybridizing probe is then identified by autoradiography, and the colony corresponding to the positive signal is isolated and further tested.

The probe can be virtually any convenient length oligonucleotide, but longer probes are more likely to identify only the specific DNA of interest. If partial amino acid sequence information is used to design a probe, then because the genetic code is degenerate, multiple codons may specify a particular amino acid. Because a positive identification in the screening procedure requires a high degree of base complementarity, the use of highly degenerate nucleotides such as those coding for Leu, Arg, and Ser, which each have six codons, should be minimized. The frequency of codon usage is not necessarily random in a particular species, and tables are available that list the frequency of codon usage for various organisms (18). Thus, peptide sequences containing tryptophan or methionine, which each have only one unique codon, are excellent choices for inclusion in a probe. Where it is not possible to eliminate a highly degenerate residue from the probe, a mixture of degenerate oligonucleotides can be synthesized simultaneously and used in screening.

2.3.3. Characterizing the Clone

After a plasmid containing the cDNA is identified by hybridization techniques, the DNA sequence is determined. If only partial protein sequence was used for screening, the complete coding sequence can be deduced by identifying the translation start signal, ATG, which codes for an initiating methionine (see Table 2), and a translation stop signal. The size or amino acid composition of the gene-encoded protein can then be compared to the size or composition of the purified protein to determine if the results are consistent.

2.4. Manipulating the Recombinant DNA Clone

Cloned DNA can be expressed or the gene can be altered by deletions, insertions, or substitutions in the cloned DNA. Deletion and insertion generally involve the use of restriction endonucleases to cut the DNA at a specific site or sites and then remove the desired segment or ligate an additional segment to the gene. Single site-specific substitution is a more common practice for the study of protein structure and function. The general procedure for substituting one amino acid residue for another involves a technique termed oligonucleotide directed mutagenesis. In this procedure, a mutagenic oligonucleotide primer of 20–30 nucleotides, containing usually one to three nucleotide mismatches to change a codon, that specify the desired amino acid substitution, is incorporated into the DNA and specifically selected for further replication. Two particularly popular procedures

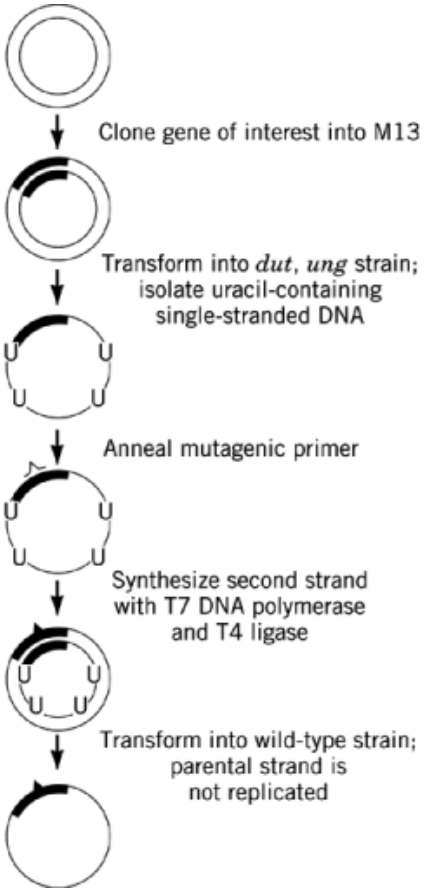

Fig. 5. Generation of mutants using single-stranded DNA. After cloning the target gene into M13, the phage is propagated in the *E. coli dut, ung* strain of *E. coli*, which results in a large percentage of uracil bases incorporated in place of thymine bases. A mutagenic primer is annealed to this DNA, and the remainder of the strand is synthesized *in vitro* with deoxyribonucleotides and DNA polymerase. The circular double-stranded DNA is then propagated in a strain of *E. coli*, which does not replicate the uracil-containing strand and therefore produces only the mutagenic DNA (21).

for this replication are use of single-stranded circular phage DNA (19), and use of the polymerase chain reaction (PCR) using double-stranded DNA (20). These methods are summarized in Figures 5 and 6, respectively.

PCR (Fig. 6) is performed in a thermocycler, which allows various reactions to occur by systematically varying the temperature. For example, denaturation of the duplex DNA to single strands (ss) is carried out at 94°C, followed by annealing of primers to the ssDNA at 45°C and extension of the primed DNA by a thermostable polymerase at 72°C. The process is then repeated 20–30 times to amplify the product. The sequences of primers 2 and 3 (Fig. 6) are complementary to the template sequence, with the exception of the base-pair mismatch(es) required to incorporate the desired mutation. Two steps are required to produce the full-length mutant gene. The first step involves two separate reaction tubes: one tube contains the DNA template and primers 1 and 2; the other tube contains the DNA template and primers 3 and 4. In the second step, the two products from the first reaction are combined into one tube, where their overlapping complementary regions (defined by primers 2 and 3 from the first reaction) base-pair to form a template for extension to the full-length DNA by the polymerase. This DNA is then replicated using only the outside primers.

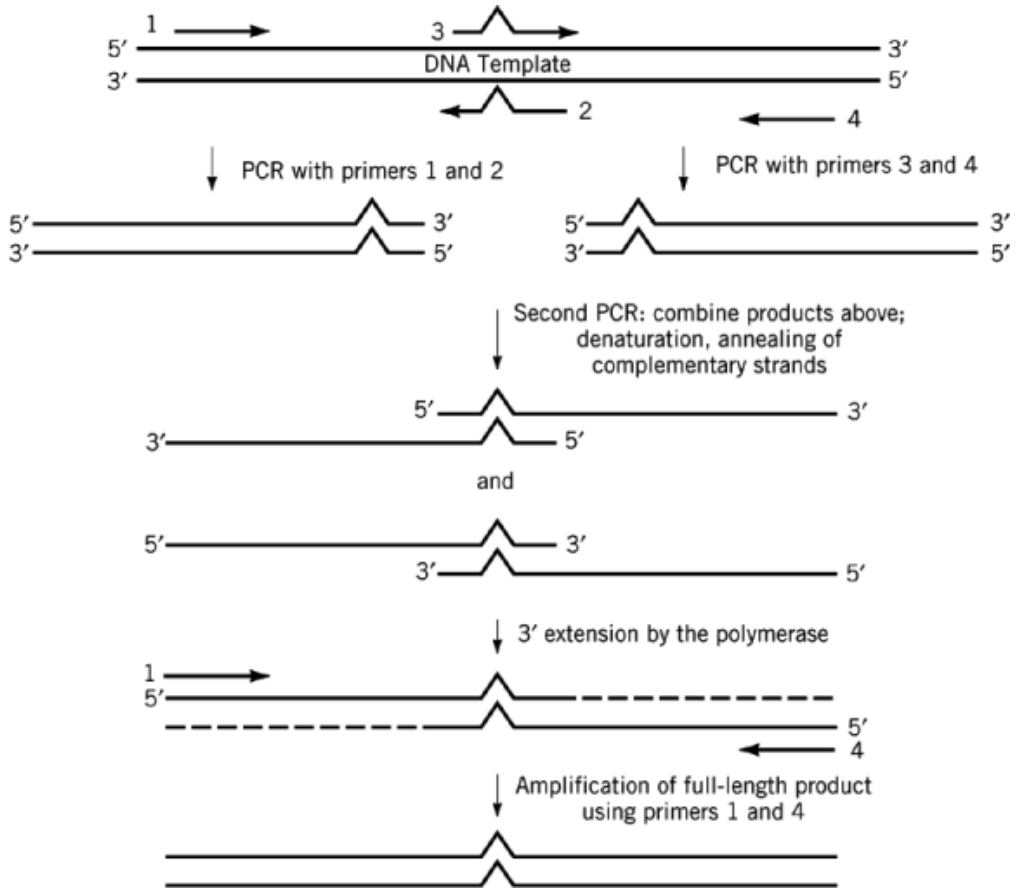

Fig. 6. Polymerase chain reaction (PCR) mediated site-directed mutagenesis. The 5′ and 3′ ends of the nucleotide strands are indicated. The four arrows surrounding the DNA template represent oligonucleotide primers 1–4. See text for discussion.

2.5. Gene Expression

2.5.1. Initial Considerations

Once the cDNA for the natural protein (wild type) or a mutant thereof is cloned, the cDNA is inserted into an appropriate vector containing a ribosome binding site and a promoter that can direct synthesis of the desired protein. The choice of vector depends on the choice of host organism in which to express the protein. Several factors must be considered. Is the gene to be expressed eukaryotic or prokaryotic? Is the protein post-translationally modified, eg, glycosylation, phosphorylation, etc, and are the modifications necessary for protein function? The original gene-encoded sequence may not adequately describe the mature protein owing to the occurrence of post-translational modifications. Is the natural protein cytoplasmic, membrane-bound, or secreted? If denatured, can the protein be easily renatured? How much protein is needed? The most common expression systems include bacteria, such as *E. coli*; mammalian cells, such as Chinese hamster ovary (CHO) and monkey kidney (COS) cells; baculovirus-infected insect cells; and yeast. Some advantages and disadvantages of these common hosts are given (22, 23).

2.5.1.1. E. coli. This bacterium has been extensively utilized as a host for protein expression (24). The organism's genome is highly flexible, accepting a great deal of manipulation. Advantages of this expression

system are that it is generally fast and inexpensive, has easy scale-up, and high level protein production is possible. A disadvantage is that the overexpressed proteins may be denatured, but this may be turned into an advantage if the denatured protein forms inclusion bodies that can be easily renatured. The cytosol of *E. coli* is such that oxidation of cysteines to form disulfide bonds does not occur. Upon lysis of the cells and exposure of the protein to the oxidizing atmosphere, disulfide bond formation can occur rapidly, sometimes with the formation of misfolded protein. Thus, this host is not always optimal for proteins containing multiple disulfide bonds. Many post-translational modifications, such as specific cleavage of propeptides and attachment of carbohydrate units (glycosylation) cannot be carried out in *E. coli*. If these post-translational modifications are important to protein function, bacterial host systems are generally a poor choice. Finally, the DNA to be transcribed cannot have introns within its sequence, because these are not processed by the bacterium.

2.5.1.2. Yeast. The advantages of expression in yeast include potentially high level production of proteins, the ability to have expressed proteins secreted into the media for ease of purification, and relatively low cost, easy scale-up. A disadvantage is that plasmid instability may be a problem which can lead to low product yield. Whereas post-translational modification occurs in yeast, proteins are quite often hyperglycosylated. This is generally a problem with expression in *Saccharomyces cerevisiae* but not for the more recently used yeast host *Pichia pastoris* (25) (see Yeasts).

2.5.1.3. Insect Cells. In this system the cDNA is inserted into the genome of an insect virus, baculovirus. Insect cells, or live insect larvae, are then infected with the virus. In this way advantage is taken of the virus's natural machinery for replication utilizing the insect cell. This is one of the best systems available for high level production of native protein having post-translational modifications similar to those seen in mammalian cells. Disadvantages of this system include lytic–batch variations, comparatively slow growth, and costly scale-up.

2.5.1.4. Mammalian. For mammalian proteins, mammalian cells offer the most natural host for expression. Problems of incorrect processing and post-translational modification are avoided using these cells. Mammalian cells are usually grown in continuous cell culture, reducing the variability in results (see Cell culture technology). Moderate-level production of native protein is possible. The procedure, however, is slow and very costly, and the level of protein expression is low. Thus large-scale production of proteins in mammalian cells is not practical. When low quantities of protein are sufficient, this system offers the several advantages described.

2.6. Purification of Expressed Proteins

Once an appropriate host has been selected and the protein has been expressed, isolation of that protein in pure form is needed. Methods for general protein purification have been described in detail in several excellent texts (26–29). These methods typically involve a combination of chromatographic separations based on physicochemical properties of the protein, such as molecular weight (size exclusion chromatography), electrostatic charge (ion exchange (qv) and chromatofocusing), or specificity of interaction with another protein or small molecule (affinity-based chromatography). Such methods are generally applicable to all proteins, whether expressed recombinantly or naturally occurring. For engineered proteins, additional purification strategies are available that take advantage of the ability to manipulate the cDNA of the target protein. Two examples are the use of polyhistidine extensions (30) and formation of fusion proteins (31). In the former, one extends the target protein cDNA to include the coding sequence for several (usually six) histidine residues. Histidines are good transition-metal chelators in their deprotonated form, ie, above pH 6. After expression and cell lysis, the lysate is applied to a transition-metal, usually nickel, affinity column. Owing to the presence of the polyhistidine extension, the target protein adheres specifically to this column whereas the majority of cellular proteins are not retained. The target protein can then be eluted from the column by lowering the pH or adding excess histidine to the mobile phase to compete for the interaction between the protein and the transition metal.

The second example involves fusing the target protein cDNA to the cDNA of another protein, making a fusion protein. The other protein is chosen because of some unique feature that makes it easy to purify.

For example, the maltose binding protein (MBP) binds tightly to an amylose-affinity column, whereas other proteins generally do not. A common strategy is to fuse the cDNA of MBP to that of a target protein in order to use amylose-affinity chromatography as a one-step purification method for the fusion protein (31). Typically, researchers also engineer in a specific protease cleavage site at the interface between the MBP sequence and that of the target protein. After amylose-affinity chromatography, the fusion protein is treated with protease to liberate the free target protein. This protein is then separated from MBP by some conventional chromatographic method, eg, size exclusion or ion-exchange chromatography.

3. Applications

3.1. Studies of Protein Stability

An understanding of the forces which stabilize and destabilize proteins is essential both to the de novo design of proteins and rational modification of existing proteins. Forces contributing to the stability of the folded protein can be covalent, such as disulfide bonds, and/or noncovalent, such as hydrophobic or electrostatic interactions. Stabilizing forces are counterbalanced by destabilizing forces such as conformational entropy and hydration, which tend to favor the unfolded form of the protein. As a result, the folded state of naturally occurring proteins has a marginal stability of only 20–60 kJ/mol (5–15 kcal/mol) (32).

Protein engineering has provided a means of assessing the contributions of various noncovalent interactions to protein stability (33–38). In general, one amino acid is replaced by another differing from the original by only one functional aspect such as size, charge, or H-bonding ability. By making a series of systematic mutations and determining the effect on protein stability, the importance of the deleted functional group can be assessed. Whereas results vary from protein to protein and depend on the details of the local environment experienced by the amino acid residue, similar results from studies of several different proteins have provided some encouraging generalizations regarding the contributions of noncovalent interactions to protein stability.

3.1.1. Measuring Protein Stability

Protein stability is usually measured quantitatively as the difference in free energy between the folded and unfolded states of the protein. These states are most commonly measured using spectroscopic techniques, such as circular dichroic spectroscopy, fluorescence (generally tryptophan fluorescence) spectroscopy, nmr spectroscopy, and absorbance spectroscopy (10). For most monomeric proteins, the two-state model of protein folding can be invoked. This model states that under equilibrium conditions, the vast majority of the protein molecules in a solution exist in either the folded (native) or unfolded (denatured) state. Any kinetic intermediates that might exist on the pathway between folded and unfolded states do not accumulate to any significant extent under equilibrium conditions (39). In other words, under any set of solution conditions, at equilibrium the entire population of protein molecules can be accounted for by the mole fraction of denatured protein, f_{d}, and the mole fraction of native protein, f_{n}, ie,

$$f_d + f_n = 1.0$$

Folded proteins can be caused to spontaneously unfold upon being exposed to chaotropic agents, such as urea or guanidine hydrochloride (Gdn), or to elevated temperature (thermal denaturation). As solution conditions are changed by addition of denaturant, the mole fraction of denatured protein increases from a minimum of zero to a maximum of 1.0 in a characteristic unfolding isotherm (Fig. 7**a**). From a plot such as Figure 7**a** one can determine the concentration of denaturant, or the temperature in the case of thermal denaturation, required to achieve half maximal unfolding, ie, where $f_d = f_n = 0.5$.

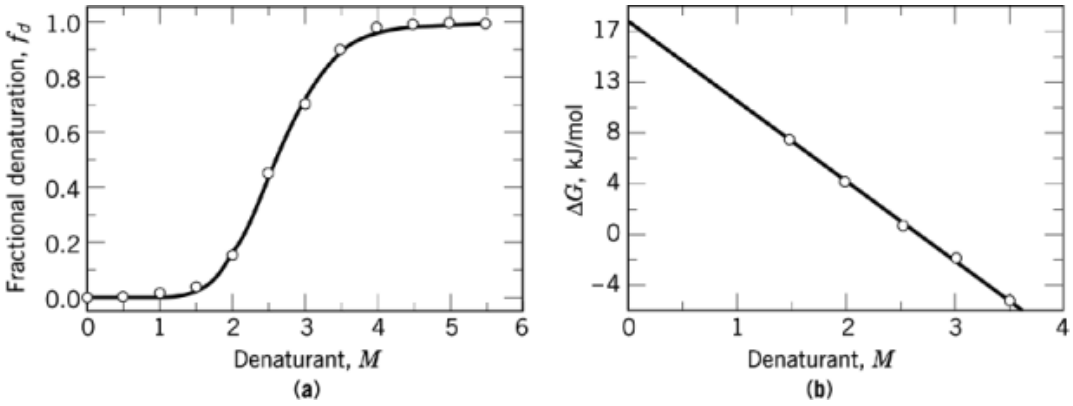

Fig. 7. Unfolding (**a**) isotherm, where the half-maximal unfolding for this protein occurs at 2.6 M denaturant; and (**b**) free energy where in the absence of denaturant, the protein has an extrapolated stability, $\Delta G^0{}_u$, of 17.6 kJ/mol (4.2 kcal/mol) as shown. To convert J to cal, divide by 4.184.

The more stable a protein is, the higher the concentration of denaturant needed to achieve half-maximal unfolding. Again invoking the two-state model, one can calculate the equilibrium constant, K, for unfolding at any point in the unfolding isotherm, and hence the Gibbs free energy of unfolding:

$$K = f_d/(1 - f_d)$$

$$\Delta G = -RT\ln(f_d/(1 - f_d))$$

In Figure 7**b**, the data are plotted as ΔG yielding a linear function. Extrapolation to zero denaturant provides a quantitative estimate of the intrinsic stability of the protein, $\Delta G^0{}_u$, which in principle is the free energy of unfolding for the protein in the absence of denaturant. Comparison of the $\Delta G^0{}_u$ values between mutant and wild-type proteins provides a quantitative means of assessing the effects of point mutations on the stability of a protein.

3.1.2. The Hydrophobic Effect

The importance of the hydrophobic core to protein stability is probed by substituting a nonpolar residue within the core with a smaller nonpolar residue. The resultant difference in free energy of unfolding can then be measured. For example, a replacement of isoleucine by valine yields information on the effect of the δ-methyl group. Further substitution by alanine or glycine studies the importance of the γ- and β-methyl groups, respectively. Studies on staphyloccocal nuclease, gene5 protein of bacteriophage $f1$, barnase, and T4 lysozyme, showed that the average effect on stability of a fully buried hydrophobic residue is 5.0 ± 1.7 kJ/mol (1.2 ± 0.4 kcal/mol) per methylene group (35, 40). There is considerable variation within these data, especially from one protein to the next. A comparison of stability data and crystal structures for several T4 lysozyme mutants (41) showed that loss of protein stability tracts linearly with the size of the cavity created by the mutation. Thus the contribution of a residue to hydrophobic stabilization depends not only on the identity of the particular substitution, but also on the size of the cavity created by that substitution.

3.1.3. Hydrogen Bonding and Salt Bridges

Historically, the importance of H-bonds to the stability of folded protein had been thought of as minor, because H-bonding interactions can also occur with water in the unfolded protein. Mutational analysis, aimed at systematically eliminating H-bonding interactions in proteins such as barnase (42), ribonuclease T1 (43), ribonuclease A, lysozyme, cytochrome c, and myoglobin (44), has been carried out by several laboratories. The results suggest that the sum of individual intramolecular H-bonds energies can contribute a significant stabilization energy, approaching that of the hydrophobic effect.

Similarly, electrostatic interactions between amino acids were thought to contribute only minimally to protein stability. Protein engineering studies have shown that this is indeed the case for salt bridges on the surface of proteins, where the high dielectric constant of the solvent water greatly reduces the electrostatic attraction (45). Disruption of buried salt bridges, on the other hand, can have a significant destabilizing effect on protein stability. For example, mutation of either partner in the buried salt bridge between His31 and Asp70 pair results in a destabilization of 13–21 kJ/mol (3–5 kcal/mol) for the protein T4 lysozyme (46).

3.1.4. Disulfides

The introduction of disulfide bonds can have various effects on protein stability. In T4 lysozyme, for example, the incorporation of some disulfides increases thermal stability; others reduce stability (47–49). Stabilization is thought to result from reduction of the conformational entropy of the unfolded state, whereas in most cases the cause of destabilization is the introduction of dihedral angle stress. In natural proteins, placement of a disulfide bond at most positions within the polypeptide chain would result in unacceptable constraint of the α-carbon chain.

3.1.5. Side-Chain Effects on Stability

The nature of the amino acid side chain can affect protein stability, not only by effecting changes in cavity sizes and hydrophobic or electrostatic interactions, but also because some side chains are not tolerable in certain secondary structure forms. For example, proline residues are known to be disruptive to α-helices. Thus, replacement of a residue within a helical region with a proline causes kinking of the helix and an overall destabilization of the folded state of the protein. Not only can a protein be destabilized by eliminating specific side-chain interactions, but replacements that cause interactions not found in the natural protein can likewise be destabilizing. For example, placement of a cysteine on the surface of a protein could lead to intermolecular disulfide bond formation between two molecules of the protein, leading to destabilization.

3.2. De Novo Designed Proteins

To further elucidate the forces which direct the protein folding pathway and stabilize the final native state, several laboratories are studying de novo (from first principles) designed proteins, that is, design of a particular protein structural motif using the level of understanding of forces believed to promote and stabilize a desired structural element. The method is then tested by evaluating the extent to which the designed protein matches the expected structure. This type of protein design is very much an iterative process, involving evaluation and subsequent modification of the design. De novo protein design has been successfully used to create proteins having all α-helix, all β-sheet, and mixed α/β secondary structures. Several excellent reviews are available (49–53).

3.2.1. α-Helical Bundles

The α-helix is the most extensively studied protein structural motif. Because α-helices form internal hydrogen bonds between the C═O of residue i and the N–H of residue $i + 4$ (see Fig. 2), the individual helix is stabilized and can exist in isolation. Individual helices can be manipulated as independent structural modules designed to associate in some predetermined manner. Often, a minimalist approach to the design of α-helices has

been taken. In this approach the goal is to obtain the desired structural motif using the simplest possible construction.

The first de novo designed, all-helical protein was an antiparallel four-helix bundle formed from a simple 16-residue sequence, α_1B, predicted to form an amphiphilic α-helix and associate as a tetramer (54). The designed sequence contains only one type of hydrophobic residue, leucine. The only charged residues are glutamate and lysine. A dimer, α_2, was then formed by linking two identical α_1B sequences by a helical hairpin to form a helix–loop–helix motif. As shown in Figure 8, the structure of the linker plays a critical role in the final structure. The short linker of a single Pro residue promoted the formation of an elongated trimer of dimers, ie, coiled coils, whereas a longer linker of Pro-Arg-Arg permitted the formation of a helix bundle. This feature of negative design was then incorporated into the final sequence, which contained the four identical helix sequences connected by the three-residue linker to form the 73-residue α_4-helix-bundle protein (55). In the α_4-helix-bundle, each of the four helices is oriented antiparallel to its two nearest neighbors and parallel to its diagonal, more-distant neighbor.

The α_4-protein is a very stable, compact protein that exhibits properties of both the native and the molten globule state (50, 56). A molten globule is a conformation having a native-like secondary structure, but less compact. This state is thought to represent an intermediate form that is accessed during protein folding. Similar to native proteins, α_4 exhibits a highly cooperative guanidinium chloride denaturation curve; it is extremely resistant to unfolding by guanidinium chloride, and its stability is sensitive to substitutions in the putative hydrophobic core of the protein. Several other experiments, however, suggest that α_4 has not achieved a complete native-like structure. Binding of the fluorescent hydrophobic dye 8-anilino-1-naphthalenesulfonic acid (ANS), which binds to the molten globule state, but much less to the native or denatured states of proteins, reveals a low order parameter, suggesting a dynamic molten globule-like hydrophobic interior. The nmr measurements of H/D exchange rates also are intermediate between molten globule and native proteins. Whereas denaturant-induced unfolding is highly cooperative, the thermal unfolding of α_4 is, again, intermediate between native and molten globule states. Thus, as predicted from the design, α_4 folds into a stable, compact state owing to hydrophobic collapse (stabilized by hydrophobic forces), but does not adopt a true native-like structure.

To produce a more native-like protein, modifications which provide specific tertiary interactions have been introduced into α_4. These include substitution of the leucine residues with amino acids possessing aromatic or β-branched side chains in order to promote specific side-chain interactions between helices (57). Metal-binding sites have also been engineered into α_4. Two variants were characterized: H3α_4 and H6α_4, which contain one and two tridentate Zn^{2+} binding sites, respectively (56). The replacement of a partially buried Leu by a His causes a less favorable free energy of folding for the two proteins in the absence of zinc, but zinc binding increases the stability significantly. The nmr spectra have shown that the compactness of the structure increases upon zinc binding for H6α_4. However, both the apo- and the metal-bound protein still bind ANS, although the order parameter for ANS bound to the Zn^{2+}-containing protein is similar to that for the binding of ANS to the natural protein, apomyoglobin.

Other approaches to de novo four-helix bundle proteins have emphasized nonrepetitive designs. One such example is the four-helix bundle protein Felix (53), a 79-residue protein which uses 19 of the 20 naturally occurring amino acids: MPEVAENFQQCLERWAKLSVGGELAHMANQAAEAILKGGNEAQLKNAQALMHEAMK-TRKYSEQLAQEFAHCAYKARASQ (see Table 1). Felix also contains a successfully designed disulfide, which links the first and fourth helices. Although Felix has no homology with any natural protein, its architecture was based on the four helix clusters of hemerythrin and cytochrome b_{562}. The designed protein has been shown to contain a very high α-helix content and is readily soluble in water.

Another approach to identifying structural determinants of four-helix bundles has employed cloning and *in vivo* expression of degenerate synthetic genes, followed by identification of proteins that form stable four-helix bundles. Based on the α-helical periodicity of 3.6 residues per turn, a sequence pattern of polar and nonpolar amino acids required to generate an amphiphilic α-helix, in which one face is predominantly

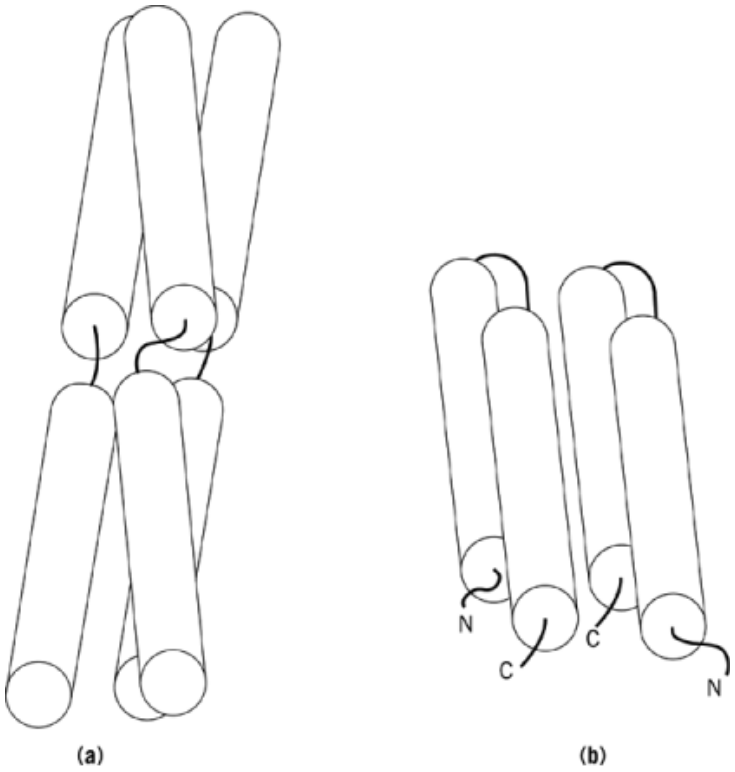

Fig. 8. De novo designed α-helical proteins. Dimers of the amphiphilic helix-forming peptide α_1B, GELEELLKKLKELLKG (see Table 1), in which the nature of the linker connecting the individual helices plays a critical role in the structure of the final protein. (**a**) Using a Pro residue as the linker, ie, α_1B-Pro-α_1B, three molecules aggregated to form a trimeric coiled-coil. (**b**) Using Pro-Arg-Arg as the linker, ie, α_1B-Pro-Arg-Arg-α_1, resulted in the formation of an antiparallel four-helix bundle.(Courtesy of S. F. Betz.)

hydrophobic and the other hydrophilic, was designed (58). Using two degenerate codons, one for polar residues and the other for nonpolar residues, a large number of synthetic genes were constructed, and the corresponding proteins were expressed. Out of 108 clones sequenced, 48 contained the desired amphipathic sequence pattern. The others contained insertions, deletions, or aberrant ligations. Twenty-nine of the 48 produced soluble proteins resistant to intracellular degradation, suggestive of compact, stable three-dimensional structures. Further characterization of three of the clones using size-exclusion chromatography and circular dichroic (cd) spectroscopy revealed proper monomeric size and helical content similar to the natural four-helix bundle protein cytochrome b_{562}. Urea denaturation curves revealed that two of the proteins possessed cooperative unfolding transitions and stabilities similar to those of natural proteins. Thus, the initial characterization of some of these clones suggests that this technique has potential value for the study of α-helical bundles.

3.2.2. β-Sheet Proteins

Successful de novo design of β-sheet proteins has been more difficult than was true of α-helical proteins (59). Because β-strands have very different H-bonding patterns compared to α-helices, several additional variables must be factored into the design. Whereas backbone H-bonding is intrasegmental within an α-helix, the backbone hydrogen bonding in a β-strand occurs between C═O and N–H groups on neighboring strands. In fact, the β-strand can form favorable interactions with neighbors in four directions, ie, left, right, up, and down.

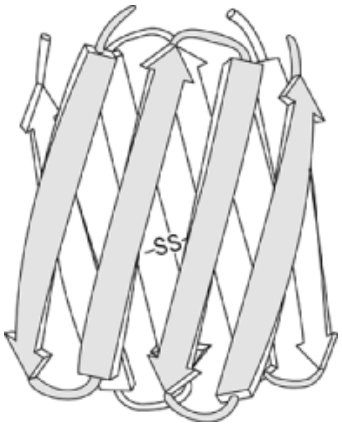

Fig. 9. A de novo designed β-sheet protein, betabellin, formed by the dimerization of two identical four-stranded β-sheets and a disulfide linking the two sheets. This model is for betabellins 9 and later progenies; the earlier betabellins contained a two-armed cross-linker connecting the sheets (51).

This neighborliness frequently causes protein aggregation and precipitation. Thus, negative design is a critical feature of β-sheet design for ensuring unfavorable contacts, where appropriate, to prevent aggregation.

Betabellin (Fig. 9) is an example of a de novo designed β-sheet protein. Betabellin contains two identical four-stranded sheets that dimerize via interaction of their hydrophobic faces to form an antiparallel β-sandwich, ie, a β-sheet, bell-shaped protein. Unlike the modular, minimalist design approach used in helical protein designs, one of the initial design criteria for the betabellins was the construction of a protein-like sequence containing normal amino acid composition, but avoiding direct homology with any natural protein. Other design criteria included secondary structure prediction, statistical preference of residues for positions in β-strands and β-turns, a pattern of alternating hydrophobic and hydrophilic residues to form two opposite faces of the sheet, and consideration of internal packing interactions (53).

Since the first betabellin was designed in the early 1980s, several progeny structures have been designed and tested, each possessing increased amounts of β-structure and improved solubility. The sequences of several of the betabellins are shown in Figure 10. This series of proteins clearly illustrates the iterative nature of de novo protein design. Betabellins 1 through 8 were formed by simultaneous synthesis of the two identical 32-residue sequences which were coupled to a two-armed cross-linker on the resin (53). Initial modifications to the design included elimination of consecutive β-branched residues that were believed to couple poorly in the peptide synthesis, and placement of a cysteine residue at position 21 to produce a disulfide across the inside of the barrel. Because these early betabellins were not very water-soluble, betabellins 7 and 8 incorporated eight additional charged residues per molecule, which slightly improved water solubility although secondary structure analysis still had to be performed on the precipitate. For betabellin 9, a significant improvement in solubility was achieved by eliminating the cross-linker and instead synthesizing the single-chain, 32-residue sequence. The cd spectroscopy performed under native conditions confirmed the β-structure, and Raman spectroscopy indicated 60–80% antiparallel β-sheet. Betabellins 12 and 14 (60) incorporated D-amino acids at the turns to promote the formation of tight β-hairpin turns. The more recently characterized β-sandwich proteins betadoublet (61) and betabellin 14D possess significantly more native-like characteristics than their predecessors. In betabellin 14, at least 17 of 32 residues are identical or conserved compared to the earlier sequences. Betadoublet has at least 15 of 32 conserved residues. Whereas the earlier species were insoluble aggregates or only marginally soluble, both of these latter structures are soluble at 10 mg/mL in aqueous buffers.

Betabellin 1:	STVTARQPNVTYSISPNTATVRLPNfTLSIG
Betabellin 2:	STLTASIPNLTYSISPNTATVKVPNYTLSIG
Betabellin 9:	HTLTASIPDLTYSIDPNTATCKVPDfTLSIGB
Betabellin 12:	HTLTASIpdLTYSINpdTATCKVpdFTLSIGA
Betabellin 14:	HSLTASIkaLTIHVQakTATCQVkaYTVHISE
Betadoublet:	TKLTATQDGLQITINDGTAKCTVDGYQVTIRS
pattern:	epnpnpnttnpnpnprrpnpnpnttnpnpnpe

Fig. 10. Sequences (see Table 1) of betabellins. In each case, only one-half of the β-sandwich is shown. The dimer is formed from identical monomeric sets of four β-strands. In the pattern sequence, e is for end, p is for polar residue, n is for nonpolar residue, and t and r are for turn residues. Lower case f is iodophenylalanine; lower case a, d, k, and p are the D-amino acid forms of alanine, aspartic acid, lysine, and proline, respectively; B is β-alanine (2, 53, 60, 61).

As for the α-helical proteins, all of the β-designed proteins still possess some characteristics of the molten globule state. Thus, whereas the hydrophobic force plays a dominant role in the folding of a protein, it is not one that is sufficient to define the unique tertiary structure of native proteins. To design a truly native-like protein de novo, specific side-chain interactions which adopt unique packing characteristics must be part of the design process. Design efforts aimed toward reaching this goal are ongoing in several laboratories.

3.3. Studies of Protein Function

For enzymes to catalyze reactions, and receptors to transport molecules and transmit signals across cellular membranes, both classes of proteins require a ligand-binding pocket, the structure of which is highly complementary to that of the ligand. It is this complementarity of structure that provides the high degree of ligand specificity characteristic of these two classes of proteins. The binding-pocket structure depends on the overall tertiary structure of the protein as well as the specific amino acid side chains that line the binding pocket. In the case of enzymes, changes in the residues within the binding pocket can greatly affect the rate or nature of the reaction catalyzed. For both enzymes and receptors, changes of amino acid residues within the binding pocket can also dramatically alter the ligand specificity. Understanding these structure–function relationships would allow the rational design of a binding pocket to alter the ligand specificity of these proteins for specific purposes, such as altered catalysis by an enzyme, or transport of unnatural molecules into cells by an engineered receptor.

3.3.1. Enzymes

Protein engineering has been used both to understand enzyme mechanism and to selectively modify enzyme function (4, 5, 62–67). Much as in protein stability studies, the role of a particular amino acid can be assessed by replacement of a residue incapable of performing the same function. An understanding of how the enzyme catalyzes a given reaction provides the basis for manipulating the activity or specificity.

Many enzymes have been the subject of protein engineering studies, including several that are important in medicine and industry, eg, lysozyme, trypsin, and cytochrome P450. Subtilisin, a bacterial serine protease used in detergents, foods, and the manufacture of leather goods, has been particularly well studied (68). This emphasis is in part owing to the wealth of structural and mechanistic information that is available for this enzyme.

3.3.1.1. Catalysis. The active site of subtilisin BPN′ contains the catalytic triad of amino acids common to the serine proteases, ie, Ser221, His64, and Asp32, as well as an oxyanion binding site, Asn155 and the main chain amide of Ser221 (Fig. 11). The catalytic turnover number, k_{cat}, is 10^9–10^{10} times greater than

the first-order spontaneous hydrolytic rate for amide substrates (69). The contribution of the residues of the catalytic triad to enzyme rate enhancement was assessed by mutating individual residues to alanine. Mutation of Ser221, His64, and Asp32 reduced k_{cat} by a factor of 2×10^6-, 2×10^6-, and 3×10^4-fold, respectively (69, 70). The fact that only small changes in K_m were observed for these mutants indicates that the reduced catalytic activity is not a result of impaired substrate binding. Even when all three residues in the catalytic triad are replaced simultaneously by alanine, the mutant retains a catalytic activity about 10^3-fold above the uncatalyzed rate. Various studies have suggested other binding interactions contribute to rate enhancement by stabilizing the transition state of the reaction. For example, elimination of the hydrogen bond to Asn155 by substitution with threonine, which is too short to form an optimal H-bond, reduces K_{cat} another 10^3-fold without significantly altering K_m (71, 72). Hence, in these mutants substrate binding is not hindered, but the rate of catalysis is reduced.

3.3.1.2. Engineering Substrate Specificity. Although the serine proteases use a common catalytic mechanism, the enzymes have a wide variety of substrate specificities. For example, the natural variant subtilisins of *B. amyloliquefaciens (subtilisin BPN′)* and *B. licheniformis (subtilisin Carlsberg)* possess very similar structures and sequences where 86 of 275 amino acids are identical, but have different catalytic efficiencies, k_{cat}/K_m, toward tetraamino acid *p*-nitroanilide substrates (67). Subtilisin Carlsberg is more active on low molecular weight synthetic substrates. The primary basis for this difference has been attributed to three residues in subtilisin BPN′ important for substrate binding: Tyr217, Glu156, and Gly169. When these residues are replaced by the corresponding residues of subtilisin Carlsberg, the substrate specificities become very similar to that of the latter species (73). Because of their efficiency in breaking down proteins, subtilisins have found use commercially in laundry cleaners. Subtilisin BPN′ is two times more efficient than subtilisin Carlsberg in laundry applications. The BPN′ Tyr217Leu mutant is 10 times more efficient than BPN′ for the hydrolysis of low molecular weight synthetic substrates, and its overall performance in laundry detergent is two times better than that of BPN′. These studies illustrate the feasibility of designing hybrid enzymes which combine the desired properties of two or more natural proteins.

Complementary steric and hydrophobic interactions have also been engineered into subtilisin BPN′ to enhance substrate specificity. In general, large hydrophobic residues are preferred in the P1 substrate position, ie, P1 refers to the first residue on the N-terminal side of the scissle bond of the substrate, 68. If, however, Gly166 in the binding pocket of the enzyme is replaced by larger residues, the volume of the binding pocket is reduced, sterically excluding large P1 side chains and enhancing the catalysis of smaller substrates (72, 74).

Specificity for a particular charged substrate can be engineered into an enzyme by replacement of residues within the enzyme-active site to achieve electrostatic complementarity between the enzyme and substrate (75). Protein engineering, when coupled with detailed structural information, is a powerful technique that can be used to alter the catalytic activity of an enzyme in a predictable fashion.

3.3.1.3. Engineering the pH Profile of Subtilisin. The activity of subtilisin BPN′ increases between pH 6 and 8 as His64 ($pK_a \sim 7.2$) is deprotonated (68). Changes in the surface charge of subtilisin have been used to shift the pH activity profile by altering the pK_a of the active site histidine. Two surface acidic residues, Asp99 and Glu156, each located more than 1 nm from the catalytic His64, were replaced with Ser and then Lys. These individual mutants and double mutants exhibited lower His64 pK_a values. The combined effects of the single mutants were cumulative in the double mutants, and in the most extreme case, the double mutant Asp99Lys/Glu156Lys had a pK_a shifted down a full pH unit (r68 and references therein). The activity of this double mutant was found to be twofold higher than the wild type enzyme at pH 8, and 10 times greater at pH 6.

3.3.1.4. Altered Enzyme Function. Manipulation of subtilisin BPN′ has been taken beyond the usual realm of protein engineering studies to include switching the primary enzyme function from peptide hydrolysis to peptide ligation. The general mechanism of serine proteases involves formation of an acyl enzyme intermediate followed by hydrolysis of this intermediate to release the peptide containing a free carboxy terminus. However, an alternative is also possible. In mixed or pure organic solvents, aminolysis, in which a primary

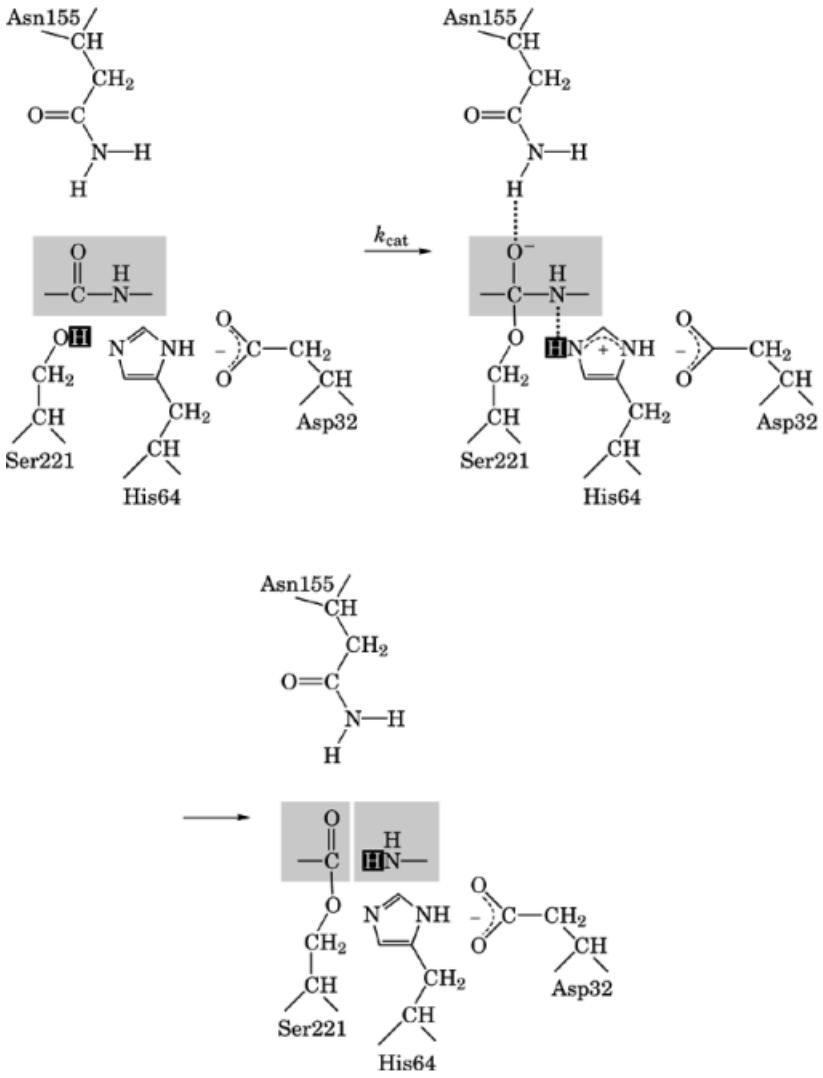

Fig. 11. Active site of the serine protease subtilisin BPN′. The amide bond of the peptide substrate is shown in gray. In the rate-limiting acylation step for peptide hydrolysis, k_{cat}, the oxygen atom on the hydroxyl of Ser221 performs a nucleophilic attack on the carbonyl of the substrate, forming a tetrahedral intermediate. His64 facilitates this nucleophilic attack of Ser221 by accepting the hydrogen atom on the hydroxyl. The positively charged His64 is in turn stabilized by the negatively charged Asp32. The substrate peptide bond is cleaved when His64 donates a proton to the substrate amide nitrogen. Deacylation (not shown) occurs by hydrolysis of the acyl enzyme intermediate in essentially the reverse of the steps above (69). Reprinted with permission from Macmillan Magazines Limited.

amine attacks the acyl enzyme intermediate, is favored over hydrolysis (76). Thus, ligation of the original peptide to another peptide is possible. However, the fact that enzymes are generally not stable in organic solvents precludes the practical use of this synthetic process. A derivative of subtilisin BPN′, in which Ser221 was chemically converted to a cysteine, was shown to favor aminolysis over hydrolysis by $>1000-$ fold (77). The Ser221Cys mutant, termed thiolsubtilisin, was shown to have greater ligase activity than wild-type subtilisin, but still retained significant amidase activity, making its use impractical for synthetic purposes. To improve

the ligase activity of thiolsubtilisin, a second mutation, Pro225Ala, was incorporated. The resulting double mutant, termed subtiligase, had a 100-fold reduced amidase activity and 10-fold greater ligase activity than the Ser221Cys mutant, yielding over 95% aminolysis without peptide hydrolysis (76).

Subtiligase has been used to sequentially ligate (couple) synthetic peptide fragments corresponding to RNase A. A full-length protein having higher yield and purity than previously reported for other synthetic processes was generated (78). Characterization of the final product showed it to be identical to commercial RNase A. The use of subtiligase offers advantages over *in vitro* transcription and translation methods, including the ability to incorporate multiple, and different, substitutions and higher yield of purified protein product. The principal disadvantage of its use is that the technology involved is limited to ligation of proteins that can be folded *in vitro*.

3.3.2. Receptors

Integral membrane receptor proteins serve critical roles in the cellular functions of transport and signal transduction. The essential and diverse roles of these receptor proteins make them excellent targets for protein engineering studies. Protein engineering techniques have been used to map receptor binding sites for the various molecules that interact with the receptor, such as agonist, antagonist, and intracellular mediators, and to identify regulatory regions of the receptor, such as phosphorylation sites. Protein domains, as well as specific amino acid residues, involved in receptor–ligand interactions have been identified by deletion mutagenesis, site-directed mutagenesis, and the construction of chimeric proteins (6, 79–81).

3.3.2.1. The β-Adrenergic Receptor. An excellent example of the application of protein engineering techniques to an integral membrane receptor protein is found in the β-adrenergic receptor (βAR). The β-adrenergic receptor belongs to the family of G-protein coupled receptors (GPCRs). Upon agonist binding, the receptor activates a G-protein by catalyzing the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) to give the active GTP-bound form, which then initiates various intracellular events. Initial ligand specificity is determined by the receptor. Different types of GPCRs couple to specific G-proteins, which in turn stimulate distinct intracellular effector systems. In the case of the β-adrenergic receptor, binding of a catecholamine agonist, such as the biogenic amines epinephrine and norepinephrine(qv), activates the G-protein G_s, which then stimulates the enzyme adenylate cyclase, resulting in production of cyclic adenosine monophosphate (cAMP), the increased concentration of which affects other cellular events, etc.

3.3.2.2. Models of GPCRs. The extensive biochemical and pharmacological characterization of many GPCRs has been accomplished in the absence of three-dimensional structure information. This fact stands in contrast to that relating to the various other examples described above, in which the rational modification of several soluble proteins, especially enzymes, has been guided by the known three-dimensional structure. Because membrane proteins are generally difficult to crystallize, structural information has been deduced from studies other than direct crystallization. Advances in genetic engineering (qv) techniques have led to the cloning and identification of several GPCRs. The availability of a large database of protein sequences has permitted identification of common structural motifs and conserved residues that may serve essential functional or structural roles. Several models of GPCR structure have been developed (82), all having seven highly conserved hydrophobic regions, proposed to be membrane-spanning α-helices (Fig. 12). These are connected by more divergent hydrophilic regions that form alternating extracellular and intracellular loops. The seven helices are believed to form a transmembrane cluster which creates the ligand binding pocket.

3.3.2.3. Site-Directed Mutagenesis Studies. The results of several initial protein engineering studies on the beta-2 adrenergic receptor (β_2AR) that involved the construction and analysis of deletion mutants suggested that ligand binding occurs within the hydrophobic core. Therefore, several site-directed mutagenesis studies of transmembrane residues were undertaken. Adrenergic agonists and antagonists are biogenic amines having protonatable nitrogens that are critical for receptor binding. Thus all acidic residues in the transmembrane region of the receptor were mutated (83, 84). Two important aspartic acid residues, Asp113 and Asp79, were

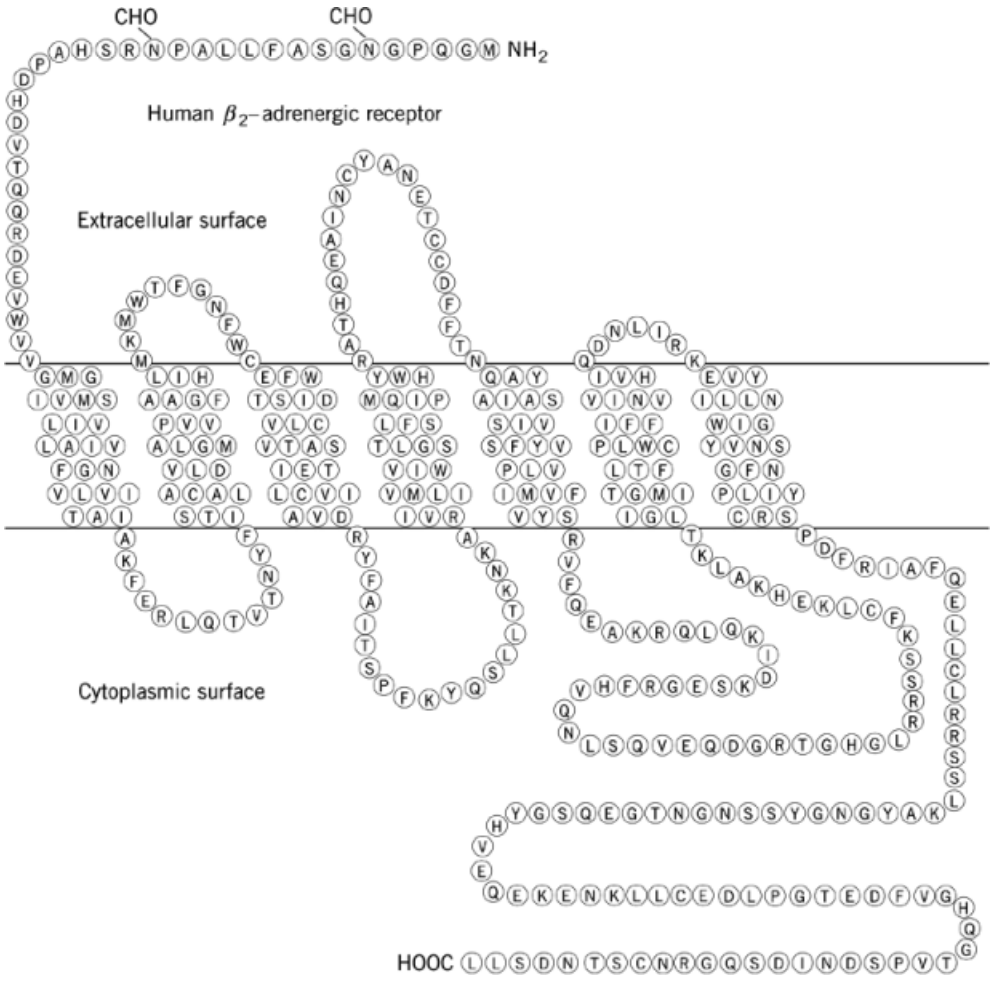

Fig. 12. Model for the β-adenergic receptor. It is proposed that the receptor possesses seven hydrophobic helices that span the plasma membrane and are connected by alternating extracellular and intracellular loops (79). The site of glycosylation is represented as CHO. Reprinted with permission from Elsevier Science Ltd., Kidlington, U.K.

identified as a result of these studies. Ligand binding affinity, K_d, for a radiolabeled antagonist was determined by saturation binding isotherms. The mutant receptor in which Asp79 was replaced by an alanine, Asp79Ala-β_2AR, had a K_d for the antagonist of 60 pM, identical to the wild-type receptor. Based on the fact that agonist and antagonist binding is both mutually exclusive and competitive, the binding affinity of nonradioactive agonist was determined by a competition assay in which the concentration of agonist required to obtain 50% maximal binding (EC_{50}) to the receptor in the presence of the antagonist was determined. These experiments showed that the mutant Asp79Ala-β_2AR exhibited a 10-fold increase in K_d for agonist (isoproterenol) binding.

The radioligand binding assays could not be used to assess antagonist or agonist binding to Asp113 mutants, owing to undetectable ligand binding. Therefore, a functional assay was used in which stimulation of the enzyme adenylate cyclase was measured. Using this assay, even very low ligand-binding affinities can be detected by measuring the functional consequences of ligand binding that result in activation of adenylate cyclase. The ability of the isoproterenol to activate adenylate cyclase was assessed by determining a dose

response curve in which the rate of production of cAMP as a function of the dose of agonist was measured. The Asp79Ala-β_2AR mutant exhibited a 10-fold increase in the activation constant, K_{act}, for adenylate cyclase stimulation. Substitution of Asp113 with Glu or Asn reduced agonist affinity such that a 10^2–10^4-fold higher concentration of agonist was required for stimulation of adenylate cyclase. These same substitutions reduced the binding affinity of the antagonist propranolol by a factor of 10^4, as assessed by inhibition of isoproterenol-stimulated adenylate cyclase activation. The fact that substitution at position 113 gives increased K_{act} for agonists and increased K_i values for antagonists suggests a similar role for Asp113 in binding these ligands. Based on these results it was concluded that Asp113, within transmembrane helix 3, is essential for high affinity agonist and antagonist binding to the receptor. The observation that binding of isoproterenol still elicited agonist activity suggests that Asp113 is essential for ligand binding, but not for functional activation of the receptor (84). Compared to the Asn mutant, substitution of Asp113 by Glu resulted in affinities closer to the wild-type values, suggesting that the carboxylate moiety may contribute some binding interaction. Although the K_{act} value increased by 2–4 orders of magnitude for Asp113Glu and Asp113Asn, the β_2 subtype ligand-specificity was maintained. Hence, the factors that determine binding selectivity for substituents on the amino groups of isoproterenol (isopropyl), epinephrine (methyl), and norepinephrine (H) do not appear to involve the Asp113 residue. Asp79 may directly interact with agonist or, more likely, maintain the conformation of the agonist-bound receptor. The existence of overlapping but distinct binding sites is supported by the observation that Asp79 mutations affect agonist but not antagonist binding. Asp79 is widely conserved in GPCRs, but Asp113 is present only in GPCRs that bind cationic amines as ligands. The role of Asp113 in determining the specificity of ligand binding was further emphasized by the demonstration (85) that replacement of the residue by a Ser converts the ligand specificity from catecholamines to catechol ketones and esters. Serine contains a hydroxyl side chain capable of hydrogen bonding to the carbonyls of ketones and esters.

Further elucidation of binding interactions which differentiate agonists from antagonists was accomplished by replacing putative hydrogen-bonding residues in the transmembrane region with alanines. These would not productively interact with the hydroxyls of the catechol (86). Pharmacophore mapping studies performed in the late 1970s suggested that hydrogen bonding interactions with the catechol hydroxyl groups were essential for agonist activation of β_2AR. Two serines, Ser204 and Ser207, which are conserved among catecholamine-binding GPCRs, were identified as important for hydrogen-bonding to the hydroxyls of the catecholamine. Mutants of both these residues exhibited normal specificity of antagonist binding but reduced antagonist affinity, as well as reduced efficacy of agonist-induced activation. Stimulation of the mutant receptors by various ligands suggested that Ser204 hydrogen bonds to the *meta*-hydroxyl of the catechol, whereas Ser207 hydrogen bonds to the *para*-hydroxyl group of the ligand. Thus, the serines orient the catechol ring in the binding pocket by hydrogen bonding to the hydroxyl groups of the ligand (Fig. 13). The localization of agonist-specific interactions with the serine residues in transmembrane helix 5 suggested a mechanism for agonist activation of the receptor. Investigations of the receptor/G-protein coupling have revealed that the third intracellular loop, connecting transmembrane helices 5 and 6, is the region of G-protein binding (88–92). Binding of agonist might cause conformational changes in helix 5 which are transmitted to the third intracellular loop to facilitate activation of G_s, and hence overall signal transduction.

3.4. Expanded Genetic Code

The protein engineering techniques described have employed site-directed mutagenesis to elucidate the role of a particular residue in protein structure or function. A technique that uses natural protein synthesis machinery site-specifically to incorporate unnatural amino acids into the protein has also been adapted (93, 94). Thus, instead of being limited to the 20 naturally occurring amino acids specified by the genetic code, an amino acid incorporating novel steric, electronic, or spectroscopic properties can be employed permitting a more systematic variation in the properties of the mutated residue. Size, shape, acidity, nucleophilicity, hydrogen-bonding strength, and hydrophobicity can all be probed. Additionally, biophysical probes such as spin and

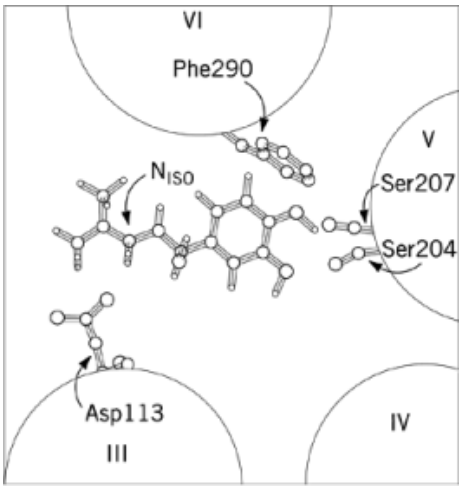

Fig. 13. Model for the ligand binding site of the β_2-adrenergic receptor (87). The view of the receptor is from the extracellular face of the plasma membrane. Four of the seven transmembrane helices are designated by Roman numerals. Arrows indicate those residues that are thought to interact with the agonist isoproterenol, on the basis of site-directed mutagenesis studies. The structure of isoproterenol in the proposed binding pocket is illustrated. The position of the amino nitrogen (N_{ISO}) is highlighted.

isotopic labels can be incorporated into a protein at the site of interest. This technique has been applied to the study of protein stability, enzymatic catalysis, and signal transduction.

The procedure for incorporating a site-specific unnatural amino acid involves the following steps. The codon for the amino acid of interest is replaced with the stop codon UAG, using standard oligonucleotide-directed mutagenesis. A suppressor tRNA that recognizes this codon is then chemically acylated using the desired unnatural amino acid. Transcription and translation of the mutagenized mutated gene is performed *in vitro* using an *E. coli* or rabbit reticulocyte extract to synthesize the mutant protein containing the unnatural amino acid at the specified position (93).

3.4.1. Examples

3.4.1.1. Protein Stability. Because it is generally accepted that the hydrophobic effect is the principal force in stabilizing globular proteins, amino acids which increase the bulk of buried hydrophobic surface area (packing density) should increase the stability of the protein. Many studies have been performed on the role of cavities in T4 lysozyme. Using natural amino acid replacements, some mutations designed to increase protein stability by filling the largest hydrophobic cavities had a slight destabilizing effect, owing to disruptive interactions with neighboring residues and the adoption of unfavorable dihedral angles (95). In contrast, the unnatural amino acids, *S*,*S*-2-amino-4-methylhexanoic acid and *S*-2-amino-3-cyclopentylpropanoic acid, designed to fill the cavity with minimal strain, increased the thermal stability of the enzyme (96). Other studies, using unnatural amino acids have probed the importance of H-bonding, and β-branching residues in the stabilization of α-helices (93, 97).

3.4.1.2. Enzyme Mechanism. Staphylococcal nuclease (SNase) accelerates the hydrolysis of phosphodiester bonds in nucleic acids (qv) some 10^{16}-fold over the uncatalyzed rate (r93 and references therein). Mutagenesis studies in which Glu43 has been replaced by Asp or Gln have shown Glu to be important for high catalytic activity. The enzyme mechanism is thought to involve base catalysis in which Glu43 acts as a general base and activates a water molecule that attacks the phosphodiester backbone of DNA. To study this mechanistic possibility further, Glu was replaced by two unnatural amino acids, homoglutamate, which is one

methylene group longer than Glu, and *S*-4-nitro-2-aminobutyric acid, which is isoelectronic and isosteric to glutamate but a much poorer base (98). Both mutants had kinetic constants similar to the wild-type enzyme. The conclusion from these studies is that Glu43 may not act as a general base but instead may serve a structural role as a bidentate H-bond acceptor, a role both unnatural substitutions could fill, to fix the conformation of the neighboring loop. Results using deletion mutagenesis of the loop (99) have suggested that this loop may be important in the product dissociation step of the catalytic mechanism.

3.4.1.3. Cellular Signal Transduction. Ras p21 is a low molecular weight GTP-binding protein that is important in cell growth and differentiation. When GTP is bound to the enzyme it is in the active signaling state and the signaling is turned off as the GTP is hydrolyzed to GDP. Point mutations that decrease the intrinsic GTPase activity of Ras or the GTPase activity stimulated by the GTPase-activating protein (GAP) are associated with approximately 30% of human cancers. Several regions of Ras, including loop L4 (switch II region), loop L2 (switch I region), and loop L1 (phosphate binding loop) have been studied by replacement with natural and unnatural amino acids (93 and references therein). For example, the conformation of loop L2 is different in the GTP and GDP bound forms. It has been proposed that a cis–trans isomerization of Pro34 may be responsible for the differences in the loop conformation. However, when Pro34 was replaced with a proline analogue that is conformationally locked in the trans state (2,4-methanoproline), no differences were observed in GTPase activity (100, 101). Therefore, the cis–trans isomerization of Pro is not essential for signal transduction.

BIBLIOGRAPHY

Cited Publications

1. A. R. Rees, M. J. E. Sternberg, and R. Wetzel, *Protein Engineering: A Practical Approach*, IRL Press at Oxford University Press, New York, 1992.
2. D. L. Oxender and C. F. Fox, eds., *Protein Engineering*, Alan R. Liss, Inc., New York, 1987.
3. C. Robinson, ed., *Trends Biotechnol.* **12**, 145–211 (1994). The entire May issue is devoted to the topic of Protein Engineering.
4. A. Fersht and G. Winter, *Trends Biochem. Sci.* **17**, 292–294 (1992); A. Recktenwald, D. Schomburg, and R. D. Schmid, *J. Biotechnol.* **28**, 1–23 (1993).
5. R. A. Copeland, *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, VCH, New York, 1996.
6. C. Chothia, *J. Mol. Biol.* **105**, 1–14 (1976).
7. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105–132 (1982).
8. M. H. Klapper, *Biochem. Biophys. Res. Commun.* **78**, 1018–1024 (1977).
9. M. O. Dayhoff, R. M. Schwartz, and B. C. Orcutt, in M. Dayhoff, ed., *Atlas of Protein Sequence and Structure*, Vol. **5**, National Biomedical Research Foundation, Washington, D.C., 1978, 345–352.
10. R. A. Copeland, *Methods for Protein Analysis: A Practical Guide to Laboratory Protocols*, Chapman & Hall, New York, 1994.
11. W. H. J. Ward, D. Timms, and A. R. Fersht, *Trends Pharmacol. Sci.* **11**, 280–284 (1990).
12. D. E. McRee, *Practical Protein Crystallography*, Academic Press, Inc., New York, 1993.
13. O. Jardetzky and G. C. K. Roberts, *NMR in Molecular Biology*, Academic Press, Inc., Orlando, Fla., 1981.
14. R. G. Krishna and F. Wold, in A. Meister, ed., *Advances in Enzymology and Related Areas of Molecular Biology*, Vol. **67**, John Wiley & Sons, Inc., New York, 1993, 265–298.
15. T. E. Hugli, *Techniques in Protein Chemistry*, Academic Press, Inc., San Diego, Calif., 1989.
16. J. D. Watson, M. Gilman, J. Witkowski, and M. Zoller, *Recombinant DNA*, 2nd ed., Scientific American Books, distributed by W. H. Freeman and Co., New York, 1992.

17. J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning*, 2nd ed., Cold Spring Harbor Press, C old Spring Harbor, N.Y., 1989.
18. F. M. Ausubel and co-workers, eds., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., New York, 1992.
19. T. A. Kunkel, K. Bebenek, and J. McClary, *Methods Enzymol.* **204**, 125–139 (1991).
20. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, *PCR Protocols: A Guide To Methods and Applications*, Academic Press, Inc., San Diego, Calif., 1990.
21. BioRad product bulletin, Lit-245 0890 90-0910a, *Muta-Gene M13 In Vitro Mutagenesis Kit*, Cat. No. 170-3580, BioRad Laboratories, Hercules, Ca., 1990.
22. *Methods in Enzymology* **185** (1990). The entire volume covers gene expression.
23. J. Hodgson, *Bio/Technol.* **11**, 887–893 (1993).
24. R. C. Hockney, *Trends Biotechnol.* **12**, 456–463 (1994).
25. J. M. Cregg, T. S. Vedvick, and W. C. Raschke, *Bio/Technol.* **11**, 905–910 (1993).
26. R. K. Scopes, *Protein Purification: Principles and Practice*, 3rd ed., Springer-Verlag, New York, 1994.
27. M. P. Deutscher, ed., *Methods in Enzymology*, Vol. **182**, *Guide to Protein Purification*, Academic Press, Inc., San Diego, Calif., 1990.
28. E. L. V. Harris and S. Angal, eds., *Protein Purification Methods: A Practical Approach*, IRL Press at Oxford University Press, Oxford, U.K., 1989.
29. E. L. V. Harris and S. Angal, eds., *Protein Purification Applications: A Practical Approach*, IRL Press at Oxford University Press, Oxford, U.K., 1990.
30. M. Uhlen and T. Miks, *Methods Enzymol.* **185**, 129–143 (1990).
31. C. V. Maina and co-workers, *Gene* **74**, 365–373 (1988).
32. P. L. Privalov, *Advan. Protein Chem.* **33**, 167–241 (1979).
33. A. R. Fersht and L. Serrano, *Curr. Opin. Struct. Biol.* **3**, 75–83 (1993).
34. A. R. Fersht, S. E. Jackson, and L. Serrano, *Phil. Trans. R. Soc. Lond. A.* **345**, 141–151 (1993).
35. B. W. Matthews, *Ann. Rev. Biochem.* **62**, 139–160 (1993).
36. B. W. Matthews, *Curr. Opin. Struct. Biol.* **3**, 589–593 (1993).
37. D. Shortle, *J. Biol. Chem.* **264**, 5315–5318 (1989).
38. P. W. Goodenough and J. A. Jenkins, *Biochem. Soc. Trans.* **19**, 655–662 (1991).
39. C. N. Pace, *CRC Crit. Rev. Biochem.* **3**, 1–43 (1975).
40. C. N. Pace, *J. Mol. Biol.* **226**, 29–35 (1992).
41. A. E. Eriksson and co-workers, *Science* **255**, 178–183 (1992).
42. L. Serrano, J. T. Kellis, P. Cann, A. Matouschek, and A. R. Fersht, *J. Mol. Biol.* **224**, 783–804 (1992).
43. B. A. Shirley, P. Stanssens, U. Hahn, and C. N. Pace, *Biochemistry* **31**, 725–732 (1992).
44. P. L. Privilov and G. I. Makhatadze, *J. Mol. Biol.* **232**, 660–679 (1993).
45. S. Dao-Pin, U. Sauer, H. Nicholson, and B. W. Matthews, *Biochemistry* **30**, 7142–7153 (1991).
46. D. E. Anderson, W. J. Becktel, and F. W. Dahlquist, *Biochemistry* **29**, 2403–2408 (1990).
47. R. Wetzel, L. J. Perry, W. A. Baase, and W. J. Becktel, *Proc. Natl. Acad. Sci.* **85**, 401–405 (1988).
48. M. Matsumara and B. W. Matthews, *Methods Enzymol.* **202**, 336–356 (1991).
49. C. Sander, *Trends Biotechnol.* **12**, 163–167 (1994).
50. S. F. Betz, D. P. Raleigh, and W. F. DeGrado, *Curr. Opin. Struct. Biol.* **3**, 601–610 (1993).
51. J. S. Richardson and co-workers, *Biophys. J.* **63**, 1186–1209 (1992).
52. W. F. DeGrado, D. P. Raleigh, and T. Handel, *Curr. Opin. Struct. Biol.* **1**, 984–993 (1991).
53. J. S. Richardson and D. C. Richardson, *Trends Biochem. Sci.* **14**, 304–309 (1989).
54. S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.* **109**, 6751–6758 (1987).
55. W. F. DeGrado, Z. R. Wasserman, and J. D. Lear, *Science* **24**, 622–628 (1989).
56. T. M. Handel, S. A. Williams, and W. F. DeGrado, *Science* **261**, 879–885 (1993).
57. D. P. Raleigh and W. F. DeGrado, *J. Am. Chem. Soc.* **114**, 10079–10081 (1992).
58. S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, and M. H. Hecht, *Science* **262**, 1680–1685 (1993).
59. M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **91**, 8729–8730 (1994).
60. Y. Yan and B. W. Erickson, *Protein Science* **3**, 1069–1073 (1994).
61. T. P. Quinn, N. B. Tweedy, R. W. Williams, J. S. Richardson, and D. C. Richardson, *Proc. Natl. Acad. Sci. USA* **91**, 8747–8751 (1994).

62. L. Hedstrom, *Curr. Opin. Struct. Biol.* **4**, 608–611 (1994).
63. J. A. Gerlt, *Curr. Opin. Struct. Biol.* **4**, 593–600 (1994).
64. A. S. El Hawrani, K. M. Moreton, R. B. Sessions, A. R. Clarke, and J. J. Holbrook, *Trends Biotechnol.* **12**, 207–211 (1994).
65. C. Wilson and D. A. Agard, *Curr. Opin. Struct. Biol.* **1**, 617–623 (1991).
66. W. M. Atkins and S. G. Sligar, *Curr. Opin. Struct. Biol.* **1**, 611–616 (1991).
67. D. A. Estell, *J. Biotechnol.* **28**, 25–30 (1993).
68. J. A. Wells and D. A. Estell, *Trends Biochem. Sci.* **13**, 291–297 (1988).
69. P. Carter and J. A. Wells, *Nature* **332**, 564–568 (1988).
70. P. Carter and J. A. Wells, *Science* **237**, 394–399 (1987).
71. J. A. Wells, B. C. Cunningham, T. P. Graycar, and D. A. Estell, *Philos. Trans. R. Soc. London A* **317**, 415–423 (1986).
72. P. Bryan, M. W. Pantoliano, S. G. Quill, H. Y. Hsiao, and T. Poulos, *Proc. Natl. Acad. Sci. USA* **83**, 3743–3745 (1986).
73. J. A. Wells, B. C. Cunningham, T. P. Graycar, and D. A. Estell, *Proc. Natl. Acad. Sci.* **84**, 5167–5171 (1987).
74. D. A. Estell and co-workers, *Science* **233**, 659–663 (1986).
75. J. A. Wells, D. B. Powers, R. R. Bott, T. P. Graycar, and D. A. Estell, *Proc. Natl. Acad. Sci. USA* **84**, 1219–1223 (1987).
76. L. Abrahmsen and co-workers, *Biochemistry* **30**, 4151–4159 (1991).
77. K. Nakatsuka, T. Sasaki, and E. T. Kaiser, *J. Am. Chem. Soc.* **109**, 3808–3810 (1987).
78. D. Y. Jackson and co-workers, *Science* **266**, 243–247 (1994).
79. R. J. Lefkowitz, B. K. Kolbika, and M. G. Caron, *Biochem. Pharmacol.* **38**, 2941–2948 (1989).
80. T. M. Fong and C. D. Strader, *Med. Res. Rev.* **14**, 387–399 (1994).
81. T. M. Savarese and C. M. Fraser, *Biochem. J.* **283**, 1–19 (1992).
82. D. Donnelly and J. C. Findlay, *Curr. Opin. Struct. Biol.* **4**, 582–589 (1994).
83. C. D. Strader and co-workers, *Proc. Natl. Acad. Sci. USA* **84**, 4384–4388 (1987).
84. C. D. Strader and co-workers, *J. Biol. Chem.* **263**, 10267–10271 (1988).
85. C. D. Strader and co-workers, *J. Biol. Chem.* **266**, 5–8 (1991).
86. C. D. Strader, M. R. Candelore, W. S. Hill, I. S. Sigal, and R. A. F. Dixon, *J. Biol. Chem.* **264**, 13572–13578 (1989).
87. C. D. Strader, I. S. Sigal, and R. A. Dixon, *FASEB J.* **3**, 1825–1832 (1989).
88. C. D. Strader and co-workers, *J. Biol. Chem.* **262**, 16439–16443 (1987).
89. R. A. F. Dixon and co-workers, *Nature* **326**, 73–77 (1987).
90. R. A. F. Dixon and co-workers, *EMBO J.* **6**, 3269–3275 (1987).
91. B. F. O'Dowd and co-workers, *J. Biol. Chem.* **263**, 15985–15992 (1988).
92. K. A. Kjelsberg, S. Cotecchia, J. Ostrowski, M. G. Caron, and R. J. Lefkowitz, *J. Biol. Chem.* **267**, 1430 (1992).
93. V. W. Cornish and P. G. Schultz, *Curr. Opin. Struct. Biol.* **4**, 601–607 (1994).
94. J. Ellman, D. Mendel, S. Anthony-Cahill, C. J. Noren, and P. G. Schultz, *Methods Enzymol.* **202**, 310–336 (1991).
95. M. Karpusas, W. A. Baase, M. Matsumura, and B. W. Matthews, *Proc. Natl. Acad. Sci. USA* **86**, 8237–8241 (1989).
96. D. Mendel and co-workers, *Science* **256**, 1798–1802 (1992).
97. P. C. Lyu, J. C. Sherman, A. Chen, and N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA* **88**, 5317–5320 (1991).
98. J. K. Judice, T. R. Gamble, E. C. Murphy, A. M. de Vos, and P. G. Schultz, *Science* **261**, 1578–1581 (1993).
99. S. P. Hale, L. B. Poole, and J. A. Gerlt, *Biochemistry* **32**, 7479–7487 (1993).
100. H.-H. Chung, D. R. Benson, V. W. Cornish, and P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **90**, 10145–10149 (1993).
101. H.-H. Chung, D. R. Benson, and P. G. Schultz, *Science* **259**, 806–809 (1993).

JUNE P. DAVIS
The Du Pont Merck Research Laboratories
ROBERT A. COPELAND
The Du Pont Merck Research Laboratories

Related Articles

Genetic engineering; Nucleic acids; Cell culture technology; Proteins