APPLIED BIOCHEMISTRY AND BIOTECHNOLOGY 8, 423-436 (1983)

# REVIEW

# Enzyme Engineering

WILLIAM H. RASTETTER

*Genentech, Inc. and Genencor, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080*

Received June 8, 1983; Accepted August 1, 1983

## Abstract

Enzyme research and development efforts have been shaped by the tools and concepts available for enzyme production and utilization. A new phase of enzymology characterized by the production of modified protein catalysts has begun, made possible by recombinant DNA technology.

**Index Entries:** Enzyme modification; protein engineering; rDNA-directed mutagenesis; graphics molecular modeling; history and development of enzymology; engineering, enzyme; mutagenesis, rDNA-directed.

## Introduction

Enzymology has entered a new phase of research and development driven by the capabilities of recombinant DNA (rDNA) technology. The technology provides a means to produce wild-type and structurally altered proteins with a facility unprecedented by conventional protein chemistry. The design and rDNA-directed alteration of enzymes, or *enzyme engineering (1)* is the subject of the present paper.

The control of protein production afforded by rDNA technology is made possible through a detailed understanding at the molecular level of the structure of genes (DNA) and of the cellular processes involved in gene expression, *viz.*, transcription and translation (*2*). Enzyme engineering relies as well on an understanding at the molecular level of the structure and the structure/function relationships of proteins. Modern structural theory as it applies to proteins is our legacy from roughly two centuries of scientific discovery and discourse at the interface of chemistry and


0273-2289/83/0800-0423$2.80

biology. The interplay of these fields during the 19th and 20th centuries has stimulated controversy, at times heated debate, but ultimately has given a unified picture of the chemistry of life processes (*3*). Herein we take an historical glance at the development of the concepts of protein structure, then focus on the potential of rDNA technology to revolutionize the study of enzyme structure and function. The tools and concepts of rDNA technology are briefly reviewed as they apply to enzyme engineering, and early efforts in the production of altered enzymes are discussed.

## Development of Structural Theory

Few today would doubt that the mechanism of action of enzymes is determined by the same underlying principles as govern the chemical interactions of small molecules. Yet, we may lose sight of the progress of the field of enzymology in the past few decades and of the infant state of the science at the turn of the century. Alfred Naquet writing in 1867 (*4*) presented a view that reflects the rudimentary state of structural theory of his time:

> The albuminoid substances [proteins] . . . do not constitute, properly speaking, chemical species; they are organs or debris of organs whose history should belong to biology rather than to chemistry.

Scientific achievement and theory in the 19th century, nonetheless, set the stage for the development of enzymology. The new field of organic chemistry slowly provided the concepts of structure upon which our detailed understanding of proteins are built. By 1800 few organic compounds had been isolated in pure form from natural sources. Such materials presented unsolvable structural problems to the chemist whose concepts of valency accommodated only simple, ionic compounds. Berzelius (*5*) made a clear distinction between inorganic and organic compounds, since the latter group seemed to follow a more complex set of rules, retaining a "vital force" from living matter. The remarkable catalytic action of enzyme extracts (*6*) clearly helped perpetuate the vitalism theory.

Vitalism stood also at the center of the Liebig-Pasteur controversy in the second half of the 19th century. Pasteur maintained that "[fermentation is] absolutely dependent on the presence of microscopic organisms" (*7*), while Liebig (*8*), Berthelot (*9*), and others attributed fermentation to purely chemical action. Only at the close of the 19th century (1897) was the issue resolved by Buchner (*10*), who, using a cell-free extract of yeast, converted sugar into alcohol and $CO_2$:

> . . . the fermentation process does not demand so complicated an apparatus as is represented by the yeast cell. The agent responsible for [fermentation], on the contrary, is to be considered a dissolved substance, without doubt an albuminous body [a protein].

Despite Buchner's early lead, well into the 20th century little consensus was found among workers in the field on the chemical nature or mode of action of enzymes. By 1920 most of the naturally occurring α-amino acids had been identified in hydrolyzates from crystalline proteins. Yet, the scientific community was far from

adoption of a structural theory that would explain how enzymes are assembled from amino acid building blocks. In fact, the crystallization and characterization of urease as a protein by Sumner (*11*) in 1926 drew much skepticism from others who viewed the crystalline protein merely as an inert carrier of the catalytic agent. The issue was clouded by the recognition of other, nonprotein molecules closely associated with enzymes (prosthetic groups in today's terminology), and required for enzymatic activity. The identity of enzymes as proteins came conclusively from quantitative studies, such as partial proteolytic digestion and immunoprecipitation, performed by Sumner (*12*), Northrop (*13* ), and their coworkers.

The determination of protein structure would elude workers in the field until the development of chromatographic techniques of amino acid analysis and sequencing, and the development of X-ray crystallographic techniques for macromolecules. These tools as applied to proteins became available only in the 1940s and 50s. As late as 1939, Pauling and Neimann (*14*) argued for the peptide theory of protein structure.

> The polypeptide chain structure of proteins, with hydrogen bonds and other interatomic forces (weaker than those corresponding to covalent bond formation) acting between peptide chains, is compatible . . . with the chemical and physical properties of proteins.

The ordered structure of protein molecules in crystals was apparent from early X-ray diffraction photographs of pepsin crystals made by Bernal and Crowfoot (*15*) in 1934. However, 24 yr elapsed before the technique provided the first glimpse of the three-dimensional structure of a protein (whale myoglobin) (*16*). In the interim, the methods of amino acid analysis and protein sequencing developed by Stein, Moore, Edman, and others were elegantly exploited for determination of protein structure. By 1945 Sanger and coworkers had begun an investigation (*17*) of the structure of the polypeptide chains of insulin. Ten years later the structure of the active hormone (*18*), with disulfide links between chains, had been elucidated. The three-dimensional structure of insulin became available only in 1970 from X-ray diffraction work of Dorothy Crowfoot Hodgkin (*19*) (also see pepsin, above.)

Today, X-ray crystallography is a vital tool for determining protein structure and its relationship to the catalytic action of enzymes. Our understanding of the specificity and rate acceleration afforded by enzymes has come about from the detailed three-dimensional structures made available for many enzymes through diffraction techniques. Writing in 1894, Fischer explained the specificity of enzymes by analogy to the fit of a lock and key. His analogy (*20*) for the enzyme–substrate complex, coming over half a century before the first determination of the three-dimensional structure of a protein (*16*), was remarkably accurate.

## Study of Structure and Function

A systematic approach to the study of enzyme structure and function requires the means to make selected structural changes in the enzyme under study. Naturally

occurring mutant enzymes provide but a limited set of proteins for examination. Synthetic chemistry, on the other hand, is inadequate for the total synthesis of enzymes and mutant enzymes from amino acid building blocks. The chemical synthesis of even small proteins is so laborious as to rule out meaningful studies of structure and function by *de novo* analog synthesis. Synthetic studies have focused instead on chemical modification (*21*) of enzymes of natural origin and on the assembly of ingenious, smaller molecules as models for enzyme activity. Molecules such as the crown ethers (*22*), cryptands (*23*), and altered cyclodextrins (*24*) share some of the binding and catalytic properties of their larger counterparts, the enzymes.

As a class, however, the enzymes are unequaled as versatile catalysts. Their size often allows the substrate to be enveloped intimately by the protein molecule. The tight interaction achieves specificity for substrate, stereoselective reactivity (*25*) and, through tight transition-state binding, rate acceleration. The infinitely variable amino acid sequence allows subtle local variations in structure providing the varied effects of hydrophobic and hydrophilic interactions, solvation, and binding pockets for substrate, cofactors, and allosteric factors. The ability to build and study unnatural catalysts with such structural variation and complexity until recently has been limited, not by the imagination, but by the tools available to the chemist.

## Enzyme Engineering

Classically, the enzymologist has been constrained by the need to isolate enzymes from their natural sources. Newer recombinant technology allows the transfer of genes coding for enzymes from diverse sources into microbes that are easily grown in the laboratory or fermentation plant. The recombinant organism, thus, has become a biosynthetic factory for production of enzymes obtained with difficulty or in low purity from their natural sources.

The new technology is made possible, in fact, by enzymes that allow the specific cutting, polymerization, and splicing of genetic material (*2*). The structural region of a gene can be pared away from the native control region and tailored with new regulatory gene sequences (perhaps native to the host organism) allowing controlled and high-level gene expression during fermentation. Incorporation of the gene into a plasmid, a small, extrachromosomal, cyclic DNA molecule, allows for gene transfer to a new host and replication of the genetic material independent of the host chromosome. The frequency of plasmid replication may be high compared to that of the bacterial chromosome and cell division, leading to multiple plasmid copies, and higher expression of the cloned gene.

The unique power of recombinant technology over synthetic chemical techniques for protein production arises from the ability to replicate cells of identical genetic makeup. These "cloned" cells, derived from a single parent, provide a source of homogeneous, engineered DNA, and ultimately a plentiful source of recombinant protein of uniform composition. The mere physical methods of separation and purification, relied upon routinely by the chemist, cannot compete with the simplicity and effectiveness of the cloning technique.

The potential application of rDNA technology by the enzymologist, however, has a much broader scope. The chemical synthesis of oligonucleotides, coupled with the enzymatic methods of genetic engineering allows the assembly and replication of *altered* genes. These semisynthetic genes code for proteins of *altered amino acid sequence*. The implications of the methodology for the study of structural enzymology are far reaching.

Despite the opportunity offered by rDNA technology, the enzymologist is faced with a dilemma in the application of the new technology. Even for the small protein, utilization of rDNA-directed mutagenesis for the production of altered structures must be done with some guiding design principles. The number of amino acids in any enzyme is too large to entertain a comprehensive study of amino acid substitution.

X-ray crystallography, the essential tool in determining the three-dimensional shapes of enzymes, provides a structural basis to guide selective protein alteration. From a crystalline protein, X-ray diffraction techniques may provide a map of electron densities in the macromolecule. In a high resolution structure, features such as disulfides, aromatic rings, and α-helices are clearly discernible from the electron density patterns. For proteins of known amino acid sequence (*26*) a molecular model representing bond lengths and angles can be fitted to the three-dimensional map of electron density. Although this model may be an assembly of colored wire, wood, or plastic, increasingly the crystallographer has made use of computer graphics for model building. The model on the graphics screen (*27*) is easily assembled, manipulated, stored, and retrieved.

The graphics molecular model and time-honored chemical principles, developed initially from small molecule interactions and reactivities, form a good basis for selective enzyme modification. Graphics software provides much more than the wire, stick model (Dreiding model) or the plastic, space-filling model (Corey-Pauling-Koltun or CPK model). The computer will provide a representation of the backbone of the protein, with or without the amino acid side chains, as well as the solvent-contact surface or electrostatic surface of the macromolecule. At the graphics screen, in color, in perspective, and in real-time (*28*), the translation and rotation of an enzyme and its substrate can be depicted. Unequaled by the hand-held model, the graphics tool affords the viewer any perspective from outside or inside of the enzyme, and allows the viewing of the macromolecule one slice or section at a time. Equipped with a stereoviewer, the graphics screen treats the operator to the illusion of depth in a graphics representation which, as surrogate enzyme, becomes a powerful design tool.

The use of X-ray crystallography and molecular modeling in an interactive approach to enzyme engineering is depicted in Fig. 1. From a wild-type gene, the techniques of genetic engineering provide a plentiful source of the corresponding protein. The wild-type protein may be assayed to determine its utility in the marketplace, and in many cases, the wild-type protein will be directly developed into a product. Second generation products and enzymes for novel applications will be derived from the design loop that utilizes X-ray crystallography and molecular modeling. From the three-dimensional structure of an enzyme, a "blueprint" for an altered, unique enzyme can be obtained by applying the concepts of enzyme structure and mechanism of action. The novel structure, initially represented only

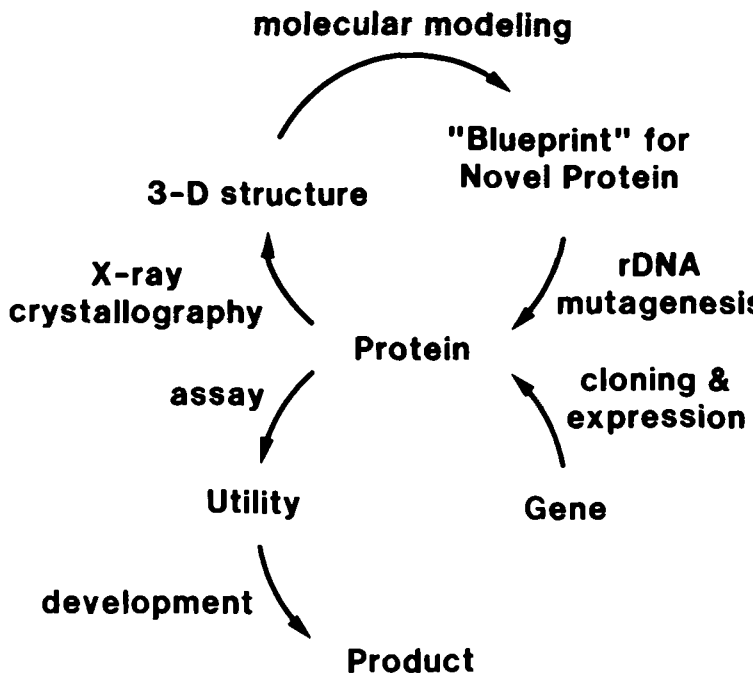


Fig. 1. The iterative steps of enzyme engineering.

on the computer graphics screen, can be tested through rDNA-directed mutagenesis of the wild-type gene and microbial biosynthesis and assay of the new protein.

The protein engineering process can and should be iterative. Crystallographic characterization of mutant proteins will provide feedback for the design process by revealing the actual 3-D structure for the altered protein. Subtle changes in structure not predicted by the graphics model may be allowed for in a second iteration of the design loop. For easily assayed mutant proteins, the iterative steps of the design process can proceed rapidly. Both the production and the crystallographic characterization of mutant proteins should be facilitated by information and material from the wild-type system, and the production and characterization of families of mutant proteins should come rapidly once the parent systems are well understood.

Examples of enzyme (protein) engineering have begun to appear in the scientific literature. Targets for rDNA-directed mutagenesis have been chosen among proteins with a known or putative relationship between structure and function. This approach has undoubtedly facilitated the interpretation of experimental results and has guided the experimentalist in the choice of site(s) for mutagenesis.

β-Lactamase is a hydrolytic enzyme of bacterial origin [encoded by the ampicillin-resistance gene ($Ap^R$) of plasmid pBR322] that deactivates the β-lactam ring of penicillin derivatives. The enzyme is biosynthesized as a precursor protein with an amino-terminus extension or signal sequence of 23 amino acids. The signal sequence is required for secretion of β-lactamase into the periplasmic space (between inner and outer membranes) of gram-negative bacteria

and is removed during the transport process. Directed mutations have provided preliminary information on structural features within the signal sequence (*29*) and the mature enzyme (*30*) that influence secretion and the removal of the signal sequence from the precursor protein.

A more detailed examination of the effect of structure and charge in the signal sequence of the prolipoprotein from *E. coli* has been reported by Inouye and coworkers (*31*).

The active site of β-lactamase has also attracted attention for structural studies by directed mutagenesis. Sigal et al. have reported (*32*) the replacement of the active site serine by cysteine, yielding a thiol-β-lactamase of reduced in vitro and in vivo activity. The activity of the mutant enzyme is rapidly eliminated by the thiophile, *p*-chloromercuribenzoate.

Richards and coworkers (*33*) have completely removed the catalytic activity of β-lactamase in a mutant protein in which the amino acids at active site positions 70 and 71 (Ser-Thr) have been reversed (Thr-Ser). Catalytic activity may depend on the presence of a primary hydroxyl group in the side chain of amino acid 70 (as in Ser) rather than a bulkier, secondary hydroxyl group (as in Thr). Random mutagenesis of the inactive Thr-Ser mutant cloned in *E. coli* yielded bacteria resistant to ampicillin indicating further mutation of the inactive enzyme to an active form. Sequencing of the gene subjected to the random mutagenesis revealed, not reversion to wild type (Ser-Thr), but a single base change in the codon for amino acid 70 giving an active enzyme with the amino acid dyad, Ser-Ser, at positions 70 and 71.

Insulin is assembled at the ribosome as a single polypeptide chain containing an amino-terminus signal peptide and an interstitial C-peptide. Maturation of insulin involves secretion and removal of the signal peptide, folding of the resultant proinsulin with formation of three disulfide bonds, and removal of the 35-amino acid C-peptide by cleavage of proinsulin at two sites. Removal of the C-peptide leaves the two separate polypeptide chains of insulin (A and B chains) connected covalently by disulfide bonds (see proinsulin in Fig. 2).

The X-ray crystallographic structure for insulin (*19*) shows that a peptide much shorter than the 35-amino acid C-peptide of human proinsulin could be made to span the gap between the carboxy-terminus of the B chain and the amino-terminus of the A chain (see Fig. 2). Such a "mini-C" analog of insulin produced by recombinant techniques is predicted to fold into a conformation closely resembling native proinsulin. Wetzel and coworkers (*34*) have constructed a gene coding for "mini-C" proinsulin in which six amino acids (Arg-Arg-Gly-Ser-Lys-Arg) replace the native C-peptide. The gene was cloned as a β-galactosidase fusion protein in *E. coli,* and the proinsulin analog obtained after cyanogen bromide cleavage and purification. Folding of the analog under conditions favoring cysteine interchange and disulfide formation in native proinsulin generated a molecule having chromatographic and radioimmunoassay behavior consistent with a proinsulin-like structure.

The X-ray crystallographic structure of tyrosyl tRNA synthetase of *Bacillus stearothermophilus* is known to ~3Å resolution and the binding sites for tyrosine and ATP have been identified. Winter and colleagues (*35*) have targeted

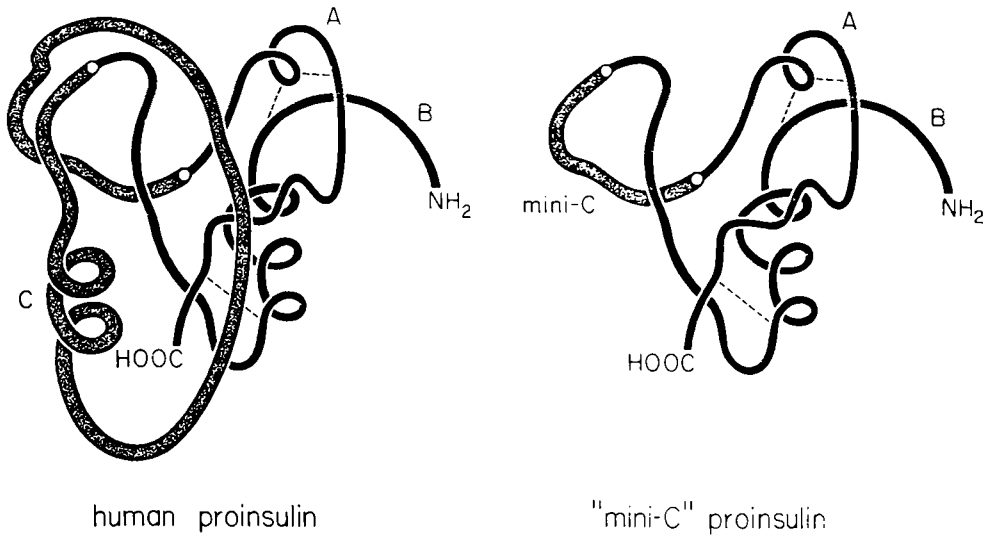


Fig. 2. Representations of native proinsulin and "mini-C" proinsulin showing the shortened connecting peptide in the engineered protein. The conformations of the A- and B-chain segments are taken from X-ray crystallographic data; the conformations of the connecting peptides were arbitrarily selected for the illustration. The overall conformation for "mini-C" insulin is assumed from the chromatographic and radioimmunoassay behavior of the refolded molecule (see text).

cysteine-35 as a site for mutagenesis based on the contact of the cysteine side chain with the 3′-hydroxyl of the sugar moiety of tyrosyl adenylate as seen in the X-ray crystal structure. Replacement of cysteine-35 by serine yields a mutant enzyme with intrinsically less activity than the wild-type. The reduction in activity is attributed largely to an altered $K_m$ for ATP.

The recent surge of papers on techniques for rDNA-directed mutagenesis attests to the rapid development and broad interest in the new methodology. An exhaustive review of the techniques for mutagenesis is beyond the scope of the present paper. A few leading papers and reviews should be noted, however.

Hutchison, Smith, and colleagues have described (*36*) a method for changing a specific nucleotide in a DNA sequence with high efficiency. A synthetic oligonucleotide complementary to the wild-type sequence except for the altered nucleotide(s) is used as a primer for the enzyme, DNA polymerase, on a circular, single stranded wild-type DNA template. Isolation of mutant sequence is achieved after biological replication, i.e., by cloning. The efficient correction of a mutation, also by use of a chemically synthesized oligonucleotide as a primer for DNA polymerase, was described by Razin et al. (*37*). Use of a chemcially synthesized oligonucleotide to introduce a mutation and as a probe to screen transformed bacterial colonies for the desired mutant has been exploited by Wallace et al. (*38*) and by others (*33*). Smith and Gillam have reviewed (*39*) the use of synthetic oligonucleotides as site-specific mutagens, and correctly suggest that "potential applications of the method are legion." Botstein, Shortle, and Rose also have reviewed (*40*) other techniques for mutagenesis of cloned genes.

As recent examples of protein engineering illustrate, rDNA techniques allow modification of proteins by amino acid substitution at any position(s) of the amino acid sequence. The power of these techniques will greatly extend the limited studies done with proteins modified by chemical and/or enzymatic techniques. The direct modification of proteins relies on the selective transformation of complex macromolecules rather than the directed and predictable synthesis of gene segments involved in recombinant mutagenesis. The former approach rarely can give the homogeneous, modified protein routinely produced by gene-splicing techniques.

The early chemical and enzymatic protein-modification experiments, nonetheless, are the conceptual forerunners of today's rDNA-directed protein engineering efforts. Two examples are noted here to illustrate the approach and the scope of direct protein modification.

The chemically modified enzyme, thiol-subtilisin (*41*), was prepared by conversion of the active site serine of subtilisin into a cysteine residue (*cf.*, thiol-β-lactamase, above). Although specific chemical reactions at single sites within proteins are not always feasible, in subtilisin the unique reactivity of the active site serine allowed its selective modification. The serine-to-cysteine transformation was accomplished via activation of the serine residue with phenylmethanesulfonyl fluoride, followed by reaction with thioacetate and deacylation of the product, acetylthiol-subtilisin. Thio-subtilisin will hydrolyze labile *p*-nitrophenyl esters, but is inactive toward normal peptide and ester substrates.

By a similar strategy, Laskowski and coworkers (*42*) in early work exploited the unique reactivity of the arginine-63–isoleucine-64 bond of soybean trypsin inhibitor toward trypsin in the direct enzymatic modification of the protein inhibitor. Modification of the residues at the cleavage site (between residues 63 and 64) allowed the activity of the inhibitor to be modulated. Although substitution of alanine at position 64 yielded a fully active trypsin inhibitor, substitution by tryptophan at position 63 gave an inhibitor for chymotrypsin. Insertion of glutamate at the cleavage site converted the trypsin inhibitor into a trypsin substrate.

Other examples and a discussion of techniques for direct protein modification have been reviewed (*21*).

Also, other reviews and comment on enzyme (protein) engineering have appeared in the literature. Ulmer (*43*) has reviewed the general approach and rationale for protein engineering, and Pabo (*44*) has discussed the general problem of protein and peptide design. On a more futuristic note, Drexler (*45*) has discussed protein engineering as an approach to the development of general capabilities for molecular manipulation, i.e., the use of proteins as tools for the microtechnologist. Among Drexler's suggestions are the assembly of molecular scale units for the storage and rapid processing of information and the *de novo* design and assembly of complex molecular machines (artificial enzymes) to direct chemical synthesis.

## Prospects for Enzyme Engineering

The study of protein structure is no longer limited by the ability to obtain altered proteins. Site specific mutagenesis allows proteins to be altered with precision.

Mutations not easily achieved by classical techniques, e.g., contiguous point mutations, are easily achieved with the newer techniques. Hybrid proteins and even *ab initio* protein design are within the scope of the technology. The field will be limited only by the ability to select the proper amino acid sequences for engineered proteins. Yet, since only a small subset of possible mutant proteins can be examined in the design and production of engineered proteins, target proteins must be carefully considered, and research must be guided by tools such as X-ray crystallography and molecular modeling.

Clearly, however, the predictive power of X-ray diffraction techniques and molecular modeling will be much greater for some amino acid substitutions than for others. Mechanism-based inactivation of enzymes will be easy to predict given a known mechanism of action for an enzyme. By contrast, changes that *improve* catalytic efficiency will be more difficult to predict given the precise alignment of functional groups likely to be necessary for efficient catalysis.

In enzyme engineering, initial successes may come from modeling changes not directly related to the mechanism of catalysis. For example, a model incorporating changes at the binding or allosteric sites of an enzyme may provide a catalyst with altered substrate specificity or sensitivity to product inhibition. Such rationally selected changes in protein structure, not selected for in the course of in vivo mutagenesis, will yield enzymes with enhanced characteristics for industrial biocatalysis.

Changes in enzyme properties that may be difficult to achieve by a rational design process and site-specific mutagenesis may be achieved through other mutagenesis techniques. Mutagenesis in continuous culture or in nature may provide active mutant enzymes, especially when the mutant provides a selective advantage to its host organism. Recombinant techniques also allow saturation mutagenesis of selected regions or sites of a gene. By these techniques, families of related enzymes are produced in a single cloning exercise. A population of bacterial colonies so produced will carry the mutant genes; each colony will harbor the gene coding for a single recombinant enzyme. Screening techniques may allow the colonies to be rapidly examined for mutant protein activity. Only the subset of colonies containing active enzymes will require further characterization. A comparison of mutant and wild-type proteins (or genes) will reveal sites that modulate activity or function. Such clues will greatly facilitate the rational, iterative design approach (see Fig. 1).

The diversity of function and range of properties of the natural enzymes reveal the broad scope of utility and variation that has evolved for the protein catalyst. The enzymologist, using the tools of rDNA technology complemented by screening and selection techniques, will build an increasingly diverse group of unqiue, human-made enzymes. The list of enzyme properties that will be targeted and enhanced or modulated for various applications is long. These properties include: substrate and cofactor specificity and affinity; reaction stereoselectivity; feedback inhibition; stability to non-aqueous solvents, heat and oxidizing media; in vivo stability; susceptibility to proteolysis; antigenicity; pH, ionic strength and temperature optima; catalytic efficiency; enzyme function and physical and chemical behavior during purification and/or immobilization or formulation. The group

of novel enzymes produced by genetic engineering will grow, initially, as limited, selected changes are made in wild-type catalysts. As we grow confident in the design and assembly of protein structrual units, hybrid enzymes bringing function and properties from two or more natrual, protein catalysts will become available. Ultimately, when protein folding can be predicted by computational methods, and proteins can be characterized more rapidly by newer crystallographic techniques, *de novo* protein design and synthesis will be possible. Protein domains or subunits will be assembled and linked at the graphics screen in a computer assisted design approach to catalyst production. The design at the graphics screen will be tested, likely domain by domain, through gene synthesis and expression, and protein isolation and characterization. These rationally conceived and deliberately assembled proteins will provide the ultimate and rigorous test of our ability to understand and manipulate protein structure, function, and activity.

How far off is *de novo* enzyme design and synthesis? Ten, fifteen years? Perhaps not that long. The answer depends largely on the development of more sophisticated knowledge of protein structure and the nature of enzyme catalysis, with an emphasis on *predictive* capability, and the development of more powerful and rapid methods for protein structural characterization. The recombinant DNA methodology for *de novo* protein synthesis exists today. A step at a time we will move away from the wild-type enzyme toward the totally synthetic protein catalyst. The new field of enzyme engineering is fertile for discovery.

## Acknowledgments

I am grateful to many colleagues at Genentech and Genencor for their improvements to my original manuscript and for their dedication and contributions to our protein engineering efforts.

## References and Notes

1. The term, *enzyme engineering*, is often used to identify the field that addresses the industrial utilization of enzymes. For our purposes the objective is the design of the catalyst, not the overall industrial process.
2. For a review, see the following: R. Wetzel (1980), "Applications of recombinant DNA technology," *Amer. Sci.* **68,** 664.
3. An in-depth review of the interplay of chemistry and biology, focusing on the period 1800–1950, is given by Joseph S. Fruton (1972), *Molecules and Life*, Wiley-Interscience, New York.
4. A. Naquet, (1867) *Principes de Chimie Fondés sur les Théories Modernes*, 2nd Ed., Savy, Paris, p. 612 (translation taken from ref. *3*).
5. See J. J. Berzelius (1813), *A View of the Progress and Present State of Animal Chemistry* (translation by G. Brunnmark), Hatchard, Johnson and Boosey, London, p. 4.

6. Amylase, separated as an alcohol precipitate of malt extract, is probably the first enzyme isolated. In 1833, the material, then named "diastase," was shown to convert starch into sugar; see A. Payen and J. F. Persoz (1833), "Mémoire su la diastase, les principaux produits de ses réactions, et leurs applications aux arts industriels," *Ann. Chem. (Phys.)* **53,** 73.
7. L. Pasteur (1878), "Première response à M. Berthelot," *Compt. Rend.* **87,** 1058.
8. J. Liebig (1839) "Uber die Erscheinung der Gährung, Fäulnis und Verwesung und ihre Ursachen, *Ann.* **30,** 250.
9. M. Berthelot (1860), "Chimie Organique Fondée sur la Synthèse," Volume II, Mallet-Bachelier, Paris, pp. 655–656.
10. E. Buchner (1897), "Alkoholische Gahrung ohne Hefezellen," *Ber. Chem. Ges.* **30,** 117.
11. (a) J. B. Sumner (1926), "The isolation and crystallization of the enzyme urease," *J. Biol. Chem.* **69,** 435. (b) J. B. Sumner (1933), "The chemical nature of enzymes," *Science* **78,** 335.
12. J. B. Sumner and G. F. Somers (1947) *Chemistry and Methods of Enzymes,* Academic Press, New York.
13. J. H. Northrop, M. Kunitz, and R. M. Herriott (1948), *Crystalline Enzymes,* Columbia University Press, New York.
14. L. Pauling and C. Niemann (1939), "The structure of proteins," *J. Amer. Chem. Soc.* **61,** 1860.
15. J. D. Bernal and D. Crowfoot (1934), "X-ray photographs of crystalline pepsin," *Nature* **133,** 794.
16. J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, and H. Wyckoff, "A three-dimensional model of the myoglobin molecule obtained by X-ray analysis," *Nature* **181,** 662.
17. F. Sanger (1945), "The free amino groups of insulin," *Biochem. J.* **39,** 507.
18. (a) A. P. Ryle, F. Sanger, L. F. Smith, and R. Kitai (1955), "The disulfide bonds of insulin," *Biochem. J.* **60,** 541, 556.
19. (a) D. C. Hodgkin (1970), "Crystal structure of insulin," *Verh. Schweiz. Naturforsch. Ges.* **150,** 93. (b) T. Blundell, G. Dodson, D. Hodgkin, and D. Mercola (1972) "Insulin: the structure in the cyrstals and its reflection in chemistry and biology," *Adv. Protein Chem.* **26,** 279.
20. E. Fischer's lock and key analogy appears in, "Einfluss der Konfiguration auf die Wirkung der Enzyme (1894), *Ber. chem. Ges.* **27,** 2992. It is clear from Fischer's later writing, however, that he never realized the true size nor the complexity of the enzymes; see E. Fischer (1923) "Untersuchungen über Aminosäuren, Polypeptide unde Proteine II, 1907–1919, Springer, Berlin, p. 8.
21. See G. E. Means and R. E. Feeney (1971), *Chemical Modification of Proteins,* Holden-Day: San Francisco.
22. See W. H. Rastetter and D. P. Phillion (1981), "Synthesis of Four Thiol-Substituted Crown Ethers," *J. Org. Chem.* **46,** 3204 and ref. 2 cited therein.
23. J. M. Lehn (1978), "The chemistry of macropolycyclic inclusion complexes," *Acct. Chem. Res.* **11,** 49.
24. (a) D. D. MacNicol, J. J. McKendrick, and D. R. Wilson (1978), "Clathrates and molecular inclusion phenomena," *Chem. Soc. Rev.* **7,** 65. (b) W. Saenger (1980), "Cyclodextrin inclusion compounds in research and industry," *Agnew. Chem. Int. Ed. Engl.* **19,** 344.
25. Producing a preponderance of a single steroisomeric product.

26. The amino acid sequence of proteins can be determined directly from the protein (see text) or more easily, when the gene encoding the amino acid sequence is available, from DNA sequencing to determine the arrangement of codons in the gene.
27. For reviews of graphics molecular modeling, see the folloing: (a) R. Langridge, T. E. Ferrin, I. D. Kuntz and M. L. Connolly (1981). "Real-time color graphics in studies of molecular interactions," *Science* **211,** 661. (b) G. Richards and V. Sackwild (1982), "Computer graphics in drug research," *Chem. Br.* **18,** 635 (c) D. A. Pensak (1983), "Modeling system cuts wasteful steps from complex R and D," *Industrial Research and Development,* p. 74. (d) R. S. Meltzer and J. T. Freeman (January/ February 1983), "Researchers use graphics modeling to search for new miracle drugs," *Computer Graphics News,* p. 3.
28. Real-time refers to the ability at the graphics screen to instantaneously and continuously rotate and translate the model, or portions of the model, in response to the operator's commands. The capability is made possible by very rapid computation of new atomic coordinates and consequent rapid changes in the graphics display which simulate motion of the model.
29. (a) A. D. Charles, A. E. Gautier, M. D. Edge, and J. R. Knowles (1982), "Targeted point mutation that creates a unique EcoRI site within the signal codons of the β-lactamase gene without altering enzyme secretion or processing," *J. Biol. Chem.* (1982) **257,** 7930. (b) D. Shortle, D. Koshland, G. M. Weinstock, and D. Botstein (1980), *Proc. Natl. Acad. Sci. USA* **77,** 5375.
30. D. Koshland and D. Botstein (1980), "Secretion of β-lactamase requires the carboxy end of the protein," *Cell* **20,** 749.
31. S. Inouye, X. Soberon, T. Franceschini, D. Nakamura, K. Itakura, and M. Inouye (1982), "Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane," *Proc. Natl. Acad. Sci. USA* **79,** 3438.
32. I. S. Sigal, B. G. Harwood, and R. Arentzen (1982), "Thiol-β-lactamase: Replacement of the active-site serine of RTEM β-lactamase by a cysteine residue," *Proc. Natl. Acad. Sci. USA* **79,** 7157.
33. (a) G. Dalbadie-McFarland, L. W. Cohen, A. D. Riggs, C. Morin, K. Itakura, and J. H. Richards (1982), "Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function," *Proc. Natl. Acad. Sci USA* **79,** 6409; (b) *Chem. Eng. News,* August 2, 1982; p. 19.
34. R. Wetzel, D. G. Kleid, R. Crea, H. L. Heyneker, D. G. Yansura, T. Hirose, A. Kraszewski, A. D. Riggs, K. Itakura, and D. V. Goeddel (1981). "Expression in *Escherichia coli* of a chemically synthesized gene for a "mini-C" analog of human proinsulin," *Gene* **16,** 63.
35. G. Winter, A. R. Fersht, A. J. Wilkinson, M. Zoeller, and M. Smith, (1982). "Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding," *Nature* **299,** 756.
36. C. A. Hutchison, III, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke, and M. Smith (1978), "Mutagenesis at a specific position in a DNA sequence," *J. Biol. Chem.* **253,** 6551.
37. A. Razin, T. Hirose, K. Itakura, and A. D. Riggs (1978), "Efficient correction of a mutation by use of chemically synthesized DNA," *Proc. Natl. Acad. Sci. USA* **75,** 4268.
38. R. B. Wallace, M. Schold, M. J. Johnson, P. Dembek, and K. Itakura (1981), "Oligonucleotide directed mutagenesis of the human β-globin gene: a general method for producing specific point mutations in cloned DNA," *Nucleic Acid Res.* **9,** 3647.

39. M. Smith and S. Gillam (1981), "Constructed mutants using synthetic oligodeoxyribonucleotides as site-specific mutagens", in *Genetic Engineering, Principles and Methods,* Vol. 3, J. K. Setlow and A. Hollaender, eds., Plenum Press, New York, p. 1.
40. D. Botstein, D. Shortle, and M. Rose (1982), "Mutagenesis of cloned genes," in *Proceedings of the IVth International Symposium on Genetics of Industrial Micoorganisms,* Y. Ikeda and T. Beppu, eds., Kodansha, Tokyo, p. 169.
41. (a) L. Polgar and M. L. Bender (1967), "The reactivity of thiol-subtilisin, an enzyme containing a synthetic functional group", *Biochemistry* **6,** 610. (b) K. E. Neet, A. Nanci, and D. E. Koshland, Jr. (1968), "Properties of thiol-subtilisin," *J. Biol. Chem.* **243,** 6392.
42. (a) R. W. Sealock and M. Laskowski, Jr. (1969), "Enzymatic replacement of the arginyl by a lysyl residue in the reactive site of soybean trypsin inhibitor", *Biochemistry* **8,** 3703. (b) T. R. Leary and M. Laskowski, Jr. (1973) "Enzymatic replacement of arginine 63 by tryptophan 63 in the reactive site of soybean trypsin inhibitor Kunitz. An intentional change from tryptic to chymotryptic specificity." *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **32,** 465. (c) D. Kowalski and M. Laskowski (1976), "Chemical-enzymatic replacement of $Ile^{64}$ in the reactive site of soybean trypsin inhibitor (Kunitz)," *Biochemistry* **15,** 1300. (d) *Idem,* "Chemical-enzymatic insertion of an amino acid residue in the reactive site of soybean trypsin inhibitor (Kunitz)," *ibid.,* p. 1309.
43. K. Ulmer (1983), "Protein engineering," *Science* **219,** 666.
44. C. Pabo (1983), "Designing proteins and peptides," *Nature* **301,** 200.
45. K. E. Drexler (1981), "Molecular engineering: an approach to the development of general capabilities for molecular manipulation," *Proc. Natl. Acad. Sci. USA* **78,** 5275.