H. Schaller, *Nucleic Acids Res.* **9**, 1919 (1981).
47. K. Strohmaier, R. Franze, K.-H. Adam, *J. Gen. Virol.* **59**, 295 (1982).
48. S. Alexander, H. Alexander, N. Green, R. A. Lerner, personal communication.
49. G. Müller, M. Shapira, R. Arnon, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 569 (1982).
50. A. M. Prince, H. Ikram, T. P. Hopp, *ibid.*, p. 579.
51. J. L. Gerin *et al.*, *ibid.*, in press.
52. D. L. Peterson, N. Natu, F. Gavilanes, *J. Biol. Chem.* **257**, 10414 (1982).
53. P. K. Bhatnagar, E. Papas, H. E. Blum, D. R. Milich, D. Nitecki, M. J. Kareis, G. N. Vyas, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4400 (1982).
54. J. H. Cox, B. Deitzschold, L. G. Schneider, *Infect. Immun.* **16**, 743 (1977).
55. A. Anilionis, W. H. Wunner, P. J. Curtis, *Nature (London)* **294**, 275 (1981).
56. B. Dietzschold, T. J. Wictor, R. MacFarlan, A. Varrichio, *J. Virol.*, in press.
57. J. Beale, *Nature (London)* **298**, 14 (1982); A. M. Q. King, B. O. Underwood, D. McCahon, J. W. I. Newman, F. Brown, *ibid.* **293**, 479 (1981).
58. L. Chedid, F. Audibert, A. Johnson, *Prog. Allergy* **25**, 63 (1978); H. Langbeheim, R. Arnon, M. Sela, *Immunology* **35**, 573 (1978); F. Audibert, M. Jolivet, L. Chedid, R. Arnon, M. Sela, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5042 (1982).
59. T. J. O'Donnell and A. J. Olson, *Computer Graphics* **15**, 133 (1981).
60. M. Connolly, unpublished results.
61. We acknowledge the efforts of our co-workers who contributed to designing the experiments and collecting the data to which we refer herein, and thank many of our colleagues for sharing their data in advance of publication. We also thank A. Olson for producing Fig. 1 and R. Ogata for comments on the manuscript. Portions of this work were supported by grants from the American Cancer Society (NP-359) and the National Institutes of Health (R01 AI 18509). This is paper No. 2824 of the Research Institute of Scripps Clinic.

Protein Engineering

Kevin M. Ulmer

Summary. The prospects for protein engineering, including the roles of x-ray crystallography, chemical synthesis of DNA, and computer modeling of protein structure and folding, are discussed. It is now possible to attempt to modify many different properties of proteins by combining information on crystal structure and protein chemistry with artificial gene synthesis. Such techniques offer the potential for altering protein structure and function in ways not possible by any other method.

In the last decade, genetic engineering technology has been developed to the point where we can now clone the gene for essentially any protein found in nature. By precise manipulation of the appropriate regulatory signals we can then produce significant quantities of that protein in bacteria. Recent advances in chemical synthesis of DNA now permit virtually unlimited genetic modification, and offer the prospect for developing protein engineering technology to create novel proteins not found in nature. By starting with the known crystal structure for a protein we would like to directly modify the gene to alter that structure in a predictable fashion, targeted to improve some functional property. At each stage we could verify the structural and functional changes that actually occurred and thereby refine and extend our predictive capability. Step by step, as we gain facility with this technique and learn the detailed rules that relate structure and function, we should be able to create proteins with novel properties which could not be achieved as effectively by any other method.

Rationale

Despite the fact that biochemists have characterized several thousand enzymes, there are only a handful that could be considered enzymes of commerce. Indeed, only a dozen enzymes have worldwide sales in excess of $10 million per year, and together they account for more than 90 percent of the total enzyme market (*1*). Frequently the limiting factor in the industrial use of an enzyme has simply been the high cost of isolating and purifying adequate amounts of the protein. Part of the solution to this problem lies with the ability of genetic engineers to greatly amplify the production of specific enzymes in microorganisms, but beyond cost there are often other limitations to the broader use of enzymes which stem from the fact that the desired industrial application is far removed from the physiological role normally played by the enzyme. In particular, industrial applications require generally robust enzymes with a long half-life under process conditions. Frequently the desired substrate or product is somewhat different from the physiological one, and often the chemical conditions for the reaction are decidedly nonphysiological, ranging to extremes of *p*H, temperature, and concentration. If enzymes are to be more widely used as industrial catalysts, we must develop methods to tailor their properties to the process of interest. The list of properties of enzymes we would like to be able to control in a predictable fashion would include the following:

1) Kinetic properties including the turnover number of the enzyme and the Michaelis constant, K_m, for a particular substrate.
2) Thermostability and temperature optimum.
3) Stability and activity in nonaqueous solvents.
4) Substrate and reaction specificity.
5) Cofactor requirements.
6) *p*H optimum.
7) Protease resistance.
8) Allosteric regulation.
9) Molecular weight and subunit structure.

The solutions to these problems have included extensive searches for the best suited naturally occurring enzyme, mutation and selection programs to enhance the native enzyme's properties, and chemical modification and immobilization to obtain a stable and functional biocatalyst. From such work we know that all of these properties can in general be improved. Specific examples of what has been achieved by these methods and how protein engineering can build on this knowledge to yield still further improvements are cited below.

It is not uncommon to observe wide variations in properties such as turnover number, K_m, molecular weight, temperature optimum, thermostability, *p*H optimum, and *p*H stability among enzymes of the same type isolated from different

Kevin M. Ulmer is director of exploratory research at Genex Corporation, Science and Technology Center, 16020 Industrial Drive, Gaithersburg, Maryland 20877.

sources. Among the glucose isomerases (E.C. 5.3.1.5) (*2*), for example, the turnover numbers range from 63 to 2151 glucose molecules converted per enzyme molecule per minute at 60°C and the K_m for glucose can differ by more than an order of magnitude (0.086 to 0.920 molar). Molecular weights vary from 52,000 to 191,000 and temperature optima vary between 50° and 90°C. Some glucose isomerases are so thermolabile that they lose all activity after exposure to 60°C for 10 minutes, while others are thermostable enough that they retain 100 percent activity after exposure to 70°C for 10 minutes. The *p*H optima differ by as much as 3.5 *p*H units, and some are stable only in the narrow range of *p*H 7 to 9 while others can tolerate the range *p*H 4 to 11. Finding the optimum combination of properties for a particular application is often a difficult task (for instance, the enzyme with the highest activity might not be the most stable) and usually results in compromise. If, instead, we could learn the structural features of each enzyme that confer a specific desirable property, we could perhaps combine these features by protein engineering techniques to create a totally new enzyme that manifests all of the desirable traits. It is difficult to imagine accomplishing this by conventional random mutagenesis techniques, but a directed approach to protein modification guided by adequate structural information should be possible.

It should also be possible to learn general rules for conferring thermostability on a protein. By examining the structures of thermophilic enzymes and comparing them with their mesophilic counterparts it has become clear that salt bridges and other electrostatic interactions confer thermostability, as do specific amino acid modifications that stabilize secondary structures and interactions between secondary structures (*3–5*). Subtle changes involving many cooperative interactions can impart significant thermostability, but the protein engineer is not necessarily limited to a subtle approach. The most thermostable enzyme may result from a combination of all these modifications, including the creation of additional disulfide bonds.

Mutagenesis and selection can often be used effectively to improve a specific property of an enzyme. For example, it is possible to isolate mutant enzymes affected in allosteric regulation which are released from feedback inhibition. The MTR 2 mutation of *Escherichia coli* anthranilate synthetase, which is insensitive to tryptophan inhibition, is such a mutant (*6*). In certain cases it has been possible to isolate mutants which have altered substrate specificity or which catalyze a different reaction from the wild-type enzyme. A mutant of xanthine dehydrogenase has been isolated, for example, which oxidizes 2-hydroxypurine at position 6 rather than position 8 (*7*).

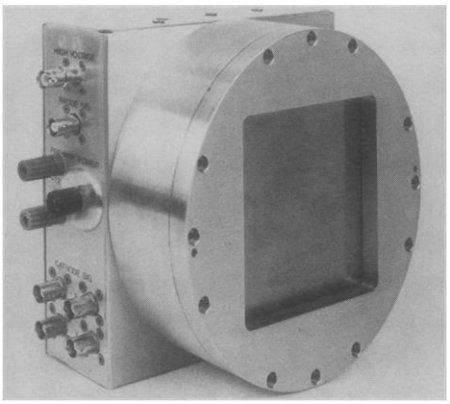

Fig. 1. Electronic position-sensitive x-ray detector. [Courtesy of Xentronics Company, Inc., Cambridge, Massachusetts]

Conventional mutagenesis techniques are generally limited to producing incremental changes in a protein. If several specific amino acid changes distributed throughout the protein are required for an observable improvement in a certain property, it will be exceedingly difficult to detect such an event in the mutant population because of the vanishingly small probability of its occurrence. If, instead, we have some guiding principles for obtaining a desired property, we can directly make whatever modifications are required by gene modification techniques.

One of the major assumptions underlying the belief that protein engineering can be successful is that proteins in general will be forgiving of attempts at modification. This view is supported by the apparent plasticity of proteins. We know from a long history of mutational studies that many amino acid changes in proteins are silent and have little or no effect on the functionality of the protein (*8*). Indeed, in many cases it is possible to isolate mutant proteins that have amino acid insertions, deletions, and substitutions and still retain normal activity, just as tryptic fragments often retain some degree of function (*9*). Many protein fusions still exhibit the activity of the two component enzymes, and in fact fusions to β-lactamase or β-galactosidase have been used as markers for studies of gene expression (*10*).

There are now enough protein structures available for a detailed comparison of enzymes from closely and distantly related organisms. It is found that there are many variations on the same theme. Proteins appear to have only a limited number of basic architectures with many subtle changes superimposed (*11*). Very similar patterns of chain folding and domain structure can arise from different amino acid sequences that show little or no homology. The immunoglobulins are a prime example of conservation of structure despite extensive differences in amino acid sequence (*12*). The natural mechanisms of evolution and gene rearrangement involve recombination of similar sequences, deletions, inversions, and duplications as well as simple point mutations. From an examination of the organization of higher eukaryotic genes it appears that functional domains of proteins may be coded in exons which are separated by introns, facilitating a building-block style of protein evolution. This is supported by recent evidence that intron-exon boundaries map at the surface of proteins (*13*). It is thus likely that many proteins will be forgiving of our initial attempts to modify their structures. Gradual changes in function and conformation should be the rule for minor changes in sequence. However, this will not always be the case. A single amino acid change (glycine to aspartic acid) in *E. coli* aspartate transcarbamylase, for example, results in loss of activity and alters the binding of catalytic and regulatory subunits (*14*). The crystals of the mutant enzyme are isomorphous with those of the native enzyme despite these extensive functional modifications. Such sensitivity to modification is likely to arise when we are dealing with critical residues in the active site, but the number of such residues should be small and thus they should be amenable to a more exhaustive analysis of the effects of modifications.

Other evidence for the likelihood of success with protein engineering comes from studies of chemical modification of enzymes. Success with semisynthetic enzymes such as flavopapain (*15*) encourages a rational approach to enzyme modification starting with crystal structure information. The proteolytic enzyme papain has been modified by specific covalent attachment of flavenoid cofactors to the unique cysteine at position 25 in the active site. From an analysis of the enzyme's structure it was predicted that such a modification would still allow room in the active site for substrate binding and would convert papain into a flavin enzyme. The flavopapain performed as expected.

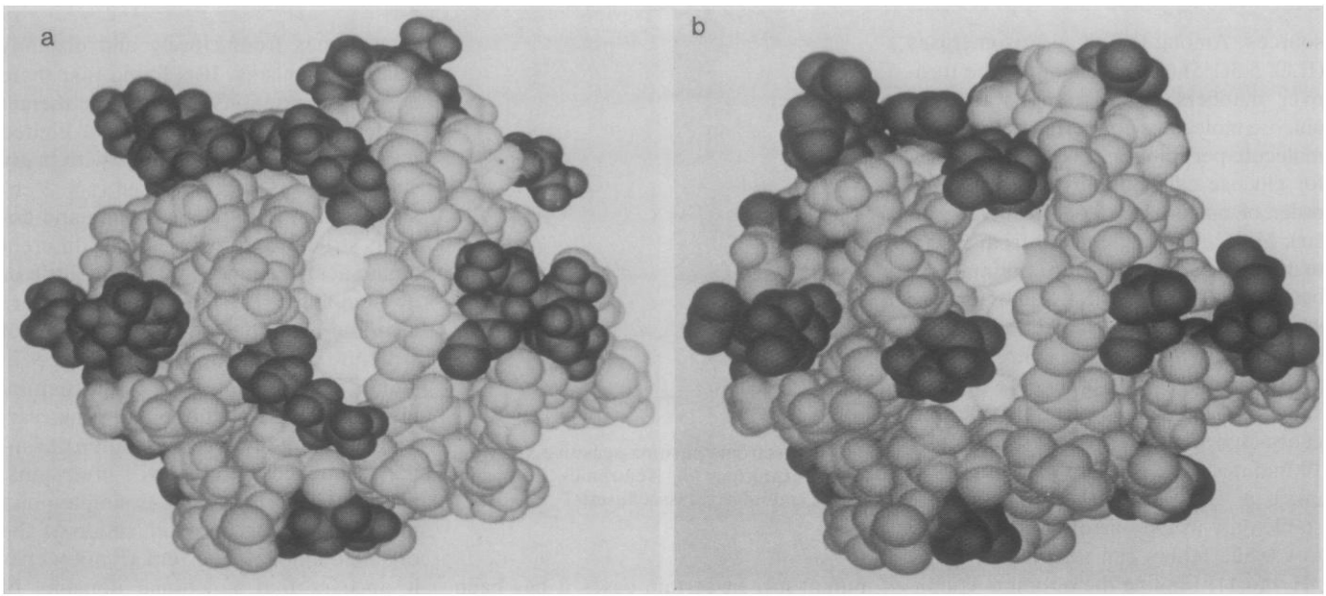

Fig. 2. Application of interactive three-dimensional computer graphics with a molecular model of tuna cytochrome c. (a) Native structure with positively charged lysine residues indicated by dark shading. (b) Lysine residues have been graphically replaced with negatively charged glutamic acid residues to simulate a protein engineering experiment that might reverse the surface charge of the protein. [Courtesy of R. J. Feldmann, National Institutes of Health, Bethesda, Maryland]

Many schemes for enzyme immobilization (*16*) also point to likely success with certain types of modifications. By more or less blindly derivatizing the surface of enzymes through the addition of polymers and other ligands (*17*), it has been possible to alter the solubility of enzymes, increase their resistance to proteases and thermal denaturation, and alter the local *p*H at the active site to advantage. All these methods are extremely crude in comparison with what should be possible starting with an accurate crystal structure for the enzyme and an artificial gene that can be specifically changed at will.

Protein Structure Determination

X-ray diffraction methods are the only techniques at present that can provide the detailed structural information at the atomic scale which will be required for protein engineering. Although protein crystallography has traditionally been a very laborious process, recent advances offer the prospect of reducing the time and effort required to solve new protein structures to 1 or 2 years. The most unpredictable aspect of the problem, which is likely to remain the rate-limiting step in the crystallographic process, is obtaining diffraction-quality crystals of the protein. Some progress has been made in recent years (*18*), but a more systematic approach with simple automated equipment could make the search for appropriate crystallization conditions more efficient. Other possibilities might include the use of zero gravity aboard the space shuttle to eliminate convective effects and improve crystallization and, once the native structure has been solved, the use of protein engineering techniques to modify the protein in order to simplify subsequent crystallizations or obtain better isomorphous derivatives.

Major advances have been made in the collection and analysis of diffraction data for proteins. Synchrotron x-ray sources are now routinely used for protein crystallography in Europe (*19*), and several facilities will soon be operational in the United States (*20*). The higher x-ray flux from such sources greatly reduces the data collection time, and the fact that the x-ray wavelength is tunable should permit phase calculation from a single isomorphous derivative by anomalous scattering techniques. The use of position-sensitive x-ray detectors (Fig. 1) in place of photographic film for recording the diffraction patterns, especially when combined with high-brilliance sources, will further reduce data collection time and simplify some of the subsequent processing steps (*21*). Better algorithms have facilitated the refinement of protein models at higher resolution (*22*), and techniques such as molecular replacement (*23*) can significantly reduce the effort required to solve related structures. The latter technique will be particularly useful for structure difference determinations, which will be required to develop protein engineering. If crystals of a modified enzyme are isomorphous with those of the native enzyme, the structural differences can be determined by a simple Fourier difference analysis, as has been done for several temperature-sensitive mutants of T4 lysozyme (*4, 5*). If the modified protein is not isomorphous, molecular replacement techniques might be used to solve the new structure with much less effort than was required for the initial structure determination. The structural differences that result from each directed modification could thus be analyzed very rapidly. It is this ability to correlate experimentally observed differences in structure with differences in functional properties that will be the key to developing predictive rules for protein engineering.

By collecting diffraction data over a range of temperatures (*24*) or by using short-pulse x-ray sources it should also be possible to learn something about the dynamic aspects of the protein structure, which are averaged out by traditional methods. It is also possible to obtain experimental data on protein dynamics by nuclear magnetic resonance (NMR) techniques. Recently, two-dimensional proton NMR techniques have been developed which may also provide detailed structural information on proteins in solution rather than in crystals (*25*). New methods permit the assignment of peaks in high-resolution NMR spectra to specific protons in the protein. A distance matrix can be constructed from such data and can then be converted to a set of three-dimensional coordinates for the molecule. So far the method has been successfully applied only to small peptides and it is not clear whether it can be extended to average-sized proteins.

Protein Modeling

Model building has also been greatly simplified through the use of sophisticated computer graphics. The protein structure can be fitted to the electron density map by simultaneously displaying both with an interactive color graphics program (*26*). Similar molecular graphics programs eliminate the need for building physical models by providing real-time, three-dimensional color representations that can be manipulated at the turn of a dial (*27*). Van der Waals surfaces for the protein can be displayed and the interaction between several molecules simulated. Interactive molecular graphics will be the design board for the protein engineer, especially when teamed with programs and superfast array processors (*28*) capable of calculating, in real time, the perturbations of a known protein structure that would result from specific modifications of the amino acid sequence (Fig. 2).

Most of the theoretical work on protein structure has been concerned with attempting to accurately predict the final three-dimensional conformation of a protein from its amino acid sequence (*29*). This is a formidable task and, although some progress has been made (*30*) in calculating the structures of small proteins such as bovine pancreatic trypsin inhibitor by using a combination of conformational energy calculations and distance constraints, the theory is not at the point where it can make significant contributions to the solution of new structures. Ultimately we hope to be able to predict structures on the basis of amino acid sequences alone, thus eliminating the need for experimental methods of structure determination. This capability will be important for the long-term success of protein engineering. Such theoretical work should benefit directly from early attempts at protein engineering, which should provide previously unobtainable experimental data to further refine algorithms or test predictive models.

For the present, however, it would be more useful to develop an accurate perturbation theory for protein structure which would allow us to calculate the effects of small changes in amino acid sequence accurately enough to eliminate the need to perform experiments for each step in the protein engineering process. The present methodology for protein structure refinement and for calculation of the structures of proteins that are homologous to other proteins of known structure would provide a useful starting point for the development of such a perturbation theory (*31*), and families of monoclonal antibodies that differ only slightly in amino acid sequence may provide a useful natural system for experimental verification of the theory. Initial attempts at protein engineering are likely to be conservative in the selection of modifications to be synthesized, but as we gain confidence in predictive modeling we will want to push the modeling to its limits in order to further extend this approach.

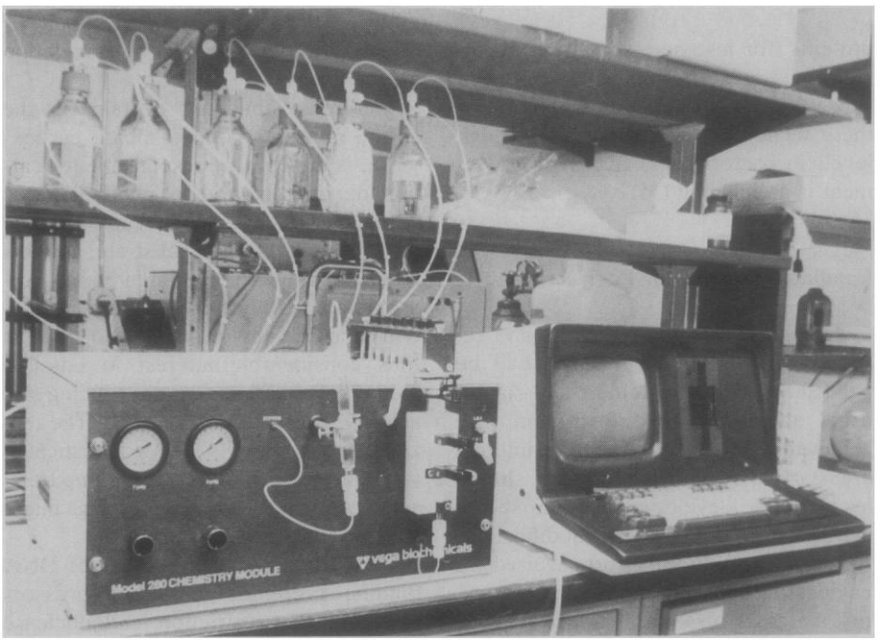

Fig. 3. Automated instrumentation for the synthesis of oligonucleotides.

Gene Modification

The technical breakthrough that makes protein engineering feasible is the ability to rapidly and inexpensively synthesize oligonucleotides of defined sequence. In the past several years the chemistry of DNA synthesis has advanced to the point where such oligonucleotides are no longer curiosities but can be considered standard laboratory reagents for the genetic engineer. This is largely due to the development of solid phase synthetic methods used in automated (Fig. 3) or semiautomated procedures (*32*). There are two general methods of gene modification with synthetic oligonucleotides.

Procedures for oligonucleotide-directed in vitro mutagenesis (*33*) are used most appropriately for making small insertions, deletions, and substitutions of nucleotides at single specific sites in cloned genes. The method is based on hybridizing a small oligonucleotide primer containing the desired nucleotide modifications to the appropriate site in a cloned gene, and then using DNA polymerase to replicate the rest of the gene, which remains unmodified. Only one modification at a time can be produced with this method, but it requires the least amount of chemically synthesized DNA and will therefore be the method of choice for most initial attempts at protein engineering. Indeed, this approach was recently used to modify the active site of tyrosyl-tRNA synthetase, whose crystal structure was known (*34*). The cysteine at position 35 was converted to serine with the predicted effect of lowering the K_m for adenosine triphosphate. This is a major step toward protein engineering.

The alternative approach is to construct a completely synthetic gene de novo, using chemically synthesized oligomers that are ligated together. This approach requires a chemical DNA synthesis capability beyond that of most laboratories at this time, but offers a number of advantages over the in vitro mutagenesis techniques. The sequence for the synthetic gene can be designed in a modular fashion which places unique restriction enzyme sites at convenient positions within the gene to facilitate subsequent modifications. Since the gene is constructed by ligation of many oligonucleotides, multiple modifications can be created in a single step by incorporating the appropriately modified oligomers into the ligation mixture. This is likely to be the gene modification method of choice when extensive changes in the amino acid sequence of the protein are required during protein engineering. The genes for several small

proteins (for instance, somatostatin, insulin, and α-interferon) have now been successfully synthesized de novo by these techniques and expressed at high levels in bacteria, and further improvement is anticipated (*35*).

Candidate Proteins for Engineering

A major investment of effort and resources in protein engineering will be needed before any commercially significant alterations to a protein are achieved. Results of academic interest are starting to appear (*34*) and should lead to the development of a set of general rules for protein modification, which will reduce the effort required for engineering subsequent proteins. To justify the costs of a major undertaking in protein engineering, the target protein should have at least some commercial potential. At present, however, we are limited in our selection of proteins by the availability of adequate structural information (*36*). Several candidates seem promising.

Immunoglobulins have been well characterized structurally (*12*) and a great deal of DNA sequence information is available (*37*). With recent advances in monoclonal antibody production, immunoglobulins have potential applications as reagents for affinity purification or as novel therapeutics, and they seem ideal candidates for protein engineering to specifically adapt them for these purposes. By creating novel gene fusions of antibodies and toxic peptides it may be possible to create targetable, cytotoxic drugs or "magic bullets" (*38*). Some success has already been achieved with the de novo design and synthesis of a toxic peptide with properties similar to those of melittin (*39*).

The α-carbon skeletons for two of the largest volume industrial enzymes, glucose isomerase (*40*) and α-amylase (*41*), have been determined. A number of properties of these enzymes have been identified (*2*) which, if improved, would greatly enhance their performance in the process for conversion of starch to high-fructose corn syrup.

Specific hydroxylation of substituted aromatic compounds is a problem of general interest in the chemical industry. The enzyme *p*-hydroxybenzoate hydroxylase, whose structure is now known (*42*), might serve as the starting point for developing an enzymatic approach to this problem. Protein engineering might be aimed at changing the substrate specificity of the enzyme as well as the position on the ring at which hydroxylation occurs (*7*).

The most abundant protein on the earth, ribulose-1,5-biphosphate carboxylase, is the enzyme responsible for carbon dioxide fixation in photosynthetic organisms. The enzyme can also use molecular oxygen as a substrate, and this results in photorespiration in plants. Approximately 50 percent of the fixed carbon is lost in this process, and there is thus considerable interest in possible methods for eliminating or reducing this activity of the enzyme (*43*). The enzyme's structure is being investigated (*44*) and the gene has been cloned and sequenced. Protein engineering might well be applied to this problem.

The structures of a number of DNA binding proteins including repressors (*45*) and the restriction endonuclease Eco RI (*46*) are receiving a great deal of attention from crystallographers and offer a number of interesting possibilities for protein engineering. It might be possible to alter the recognition specificity (*47*) of these enzymes in a predictable fashion and thus create whole new families of repressors and restriction enzymes.

Future Prospects

The ability to readily produce and analyze directed structural modifications in proteins will be of benefit in helping to solve the long-standing problem of structure-function relations in proteins. The rules learned during this academic exercise can then be applied to create novel proteins with improved properties for commercial applications. The same techniques may also assist in the development of a method for accurately predicting the three-dimensional structure of a protein from its amino acid sequence, paving the way for designing novel enzymes from first principles. Protein engineering thus represents the first major step toward a more general capability for molecular engineering which would allow us to structure matter atom by atom (*48*).

References and Notes

1. Genex Corporation estimate.
2. W. Chen, *Process Biochem.* **15**, 36 (1980).
3. M. F. Perutz, *Science* **201**, 1187 (1978); P. Argos, M. G. Rossmann, U. M. Grau, H. Zuber, G. Prank, J. D. Tratschin, *Biochemistry* **18**, 5698 (1979); M. G. Grutter, K. Rine, B. W. Matthews, *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 1009 (1979); K. Yutani, K. Ogasahara, Y. Sugino, A. Matsushiro, *Nature (London)* **267**, 274 (1977).
4. M. G. Grutter and B. W. Matthews, *J. Mol. Biol.* **154**, 525 (1978).
5. M. G. Grutter, R. B. Hawkes, B. W. Matthews, *Nature (London)* **277**, 667 (1979).
6. M. J. Pabst, J. C. Kuhn, R. L. Somerville, *J. Biol. Chem.* **248**, 901 (1973).
7. C. Scazzocchio and H. M. Sealy-Lewis, *Eur. J. Biochem.* **91**, 99 (1978).
8. J. H. Miller, in *The Operon*, J. H. Miller and W. S. Reznikoff, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1980), p. 31.
9. D. B. Wetlaufer, *Adv. Protein Chem.* **34**, 61 (1981).
10. L. Guarente, G. Lauer, T. M. Roberts, M. Ptashne, *Cell* **20**, 543 (1980).
11. M. G. Rossmann and P. Argos, *Annu. Rev. Biochem.* **50**, 497 (1981).
12. L. M. Amzel and R. J. Poljak, *ibid.* **48**, 961 (1979).
13. C. S. Craik, S. Sprang, R. Fletterick, W. J. Rutter, *Nature (London)* **299**, 180 (1982).
14. R. Kim, T. Young, H. K. Schachman, S. Kim, *J. Biol. Chem.* **256**, 4691 (1981).
15. J. T. Slama, S. R. Oruganti, E. T. Kaiser, *J. Am. Chem. Soc.* **103**, 6211 (1981); H. L. Levine, Y. Nakagawa, E. T. Kaiser, *Biochem. Biophys. Res. Commun.* **76**, 64 (1977); E. T. Kaiser, H. L. Levine, T. Otuski, H. E. Fried, R. Dupeyre, *Adv. Chem. Ser.* **191**, 35 (1980).
16. K. Mosbach, Ed., *Methods in Enzymology* (Academic Press, New York, 1976), vol. 44.
17. C. H. W. Hirs and N. Tmasheff, Eds., *ibid.* (1972), vol. 25.
18. A. McPherson, *The Preparation and Analysis of Protein Crystals* (Wiley, New York, 1982).
19. DESY at European Molecular Biology Laboratories, Hamburg, West Germany; LURE at the University of Paris-Sud, Orsay, France; DARESBURY at Daresbury, United Kingdom.
20. CHESS (Cornell High Energy Synchrotron Source), Cornell University; SSRL (Stanford Synchrotron Radiation Laboratory), Stanford University; NSLS (National Synchrotron Light Source), Brookhaven National Laboratory.
21. R. P. Phizackerley, C. W. Cork, R. C. Hamlin, C. P. Nielsen, W. Vernon, Ng. H. Xuong, V. Perez-Mendez, *Nucl. Instrum. Methods* **172**, 393 (1980); U. W. Arndt and D. J. Gilmore, *J. Appl. Crystallogr.* **12**, 1 (1979); G. E. Schulz and G. Rosenbaum, *Nucl. Instrum. Methods* **152**, 205 (1978); R. Hamlin, C. Cork, C. Nielsen, W. Vernon, Ng. H. Xuong, *Acta Crystallogr. Sect. A* **34**, 334 (1978); R. Hamlin, C. Cork, A. Howard, C. Nielsen, W. Vernon, D. Matthews, Ng. H. Xuong, *J. Appl. Crystallogr.* **14**, 85 (1981).
22. J. H. Konnert and W. A. Hendrickson, *Acta Crystallogr. Sect. A* **36**, 344 (1980).
23. T. L. Blundell and L. N. Johnson, *Protein Crystallography* (Academic Press, New York, 1976).
24. P. J. Artymuik, C. C. F. Blake, D. E. P. Grace, S. J. Oatley, D. C. Phillips, M. J. E. Sternberg, *Nature (London)* **280**, 563 (1979); H. Frauenfelder, G. A. Petsko, D. Tsernoglou, *ibid.*, p. 558; W. A. Hendrickson and J. H. Konnert, *Biophys. J.* **32**, 645 (1980).
25. K. Wuthrich, G. Wider, G. Wagner, W. Braun, *J. Mol. Biol.* **155**, 311 (1982); M. Billeter, W. Braun, K. Wuthrich, *ibid.*, p. 321; G. Wagner and K. Wuthrich, *ibid.*, p. 347; G. Wider, K. H. Lee, K. Wuthrich, *ibid.*, p. 367.
26. J. R. Miller, S. S. Abdel-Meguid, M. G. Rossmann, D. C. Anderson, *J. Appl. Crystallogr.* **14**, 94 (1981).
27. R. Langridge, T. E. Ferrin, I. D. Kuntz, M. L. Connolly, *Science* **211**, 661 (1981).
28. C. Pottle, M. S. Pottle, R. W. Tuttle, R. J. Kinch, H. A. Scheraga, *J. Comput. Chem.* **1**, 46 (1980).
29. R. F. Doolittle, *Abstr. Pap. Am. Chem. Soc.* **1979**, 44 (1979); M. J. E. Sternberg and J. M. Thornton, *Nature (London)* **271**, 15 (1978).
30. H. A. Scheraga, in *Structure and Dynamics of Proteins and Nucleic Acids*, E. Clementi and R. H. Sarma, Eds. (Adenine, Guilderland, N.Y., in press).
31. H. A. Scheraga, *Biopolymers* **20**, 1877 (1981); R. J. Feldmann, personal communication.
32. G. Alvarado-Urbina, G. M. Sathe, W.-C. Liu, M. F. Gillen, P. D. Duck, R. Bender, K. K. Ogilvie, *Science* **214**, 270 (1981); M. W. Hunkapillar and L. E. Hood, *ibid.* **219**, 650 (1983).
33. M. J. Zoller and M. Smith, in *Methods in Enzymology* (Academic Press, New York, in press).
34. G. Winter, A. R. Fersht, A. J. Wilkinson, M. Zoller, M. Smith, *Nature (London)* **299**, 756 (1982).
35. K. Itakura, T. Hirose, R. Crea, A. D. Riggs, H. L. Heyneker, F. Bolivar, H. W. Boyer, *Science* **198**, 1056 (1977); R. Crea, A. Kraszewski, T. Hirose, K. Itakura, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5765 (1978); D. V. Goeddel *et al.*, *ibid.* **76**, 106 (1979); D. C. Williams, R. M. Van Frank, W. L. Muth, J. P. Burnett, *Science* **215**,

687 (1982); M. D. Edge *et al.*, *Nature (London)* **292**, 756 (1981); J. D. Windass, C. R. Newton, J. De Maeyer-Guignard, V. E. Moore, A. F. Markham, M. D. Edge, *Nucleic Acids Res.* **10**, 6639 (1982).
36. Protein Data Bank, Brookhaven National Laboratory; F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
37. M. O. Dayhoff, R. M. Schwartz, H. R. Chen, L. T. Hunt, W. C. Barker, B. C. Orcutt, Eds., *Nucleic Acid Sequence Database* (National Biomedical Research Council, Washington, D.C., 1981).
38. T. Davies, *Nature (London)* **289**, 12 (1981); S. Olsnes, *ibid.* **290**, 84 (1981).
39. W. F. DeGrado, F. J. Kezdy, E. T. Kaiser, *J. Am. Chem. Soc.* **103**, 679 (1981).
40. H. M. Berman, B. H. Rubin, H. L. Carrell, J. P. Glusker, *J. Biol. Chem.* **249**, 3983 (1974); H. L. Carrell, *Abstr. Am. Crystallogr. Assoc.* **10**, 35 (1982).
41. Y. Matsuura *et al.*, *J. Biochem. (Tokyo)* **87**, 1555 (1980).
42. R. K. Wierenga, R. J. DeJong, K. H. Kalk, W. G. J. Hol, J. Drenth, *J. Mol. Biol.* **131**, 55 (1979).
43. H. W. Siegelman and G. Hind, Eds., *Photosynthetic Carbon Assimilation* (Plenum, New York, 1978); C. R. Somerville and W. L. Ogren, *Trends Biochem. Sci.* **7**, 171 (1982).
44. S. Johal, D. P. Bourque, W. W. Smith, S. W. Suh, D. Eisenberg, *J. Biol. Chem.* **255**, 8873 (1980).
45. C. O. Pabo, W. Krovatin, A. Jeffrey, R. T. Sauer, *Nature (London)* **298**, 441 (1982); C. O. Pabo and M. Lewis, *ibid.* p. 443; D. B. McKay and T. A. Steitz, *ibid.* **290**, 744 (1981); W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. W. Matthews, *ibid.*, p. 754; D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, B. W. Matthews, *ibid.* **298**, 718 (1982).
46. T. Young, P. Modrich, A. Beth, E. Jay, S. Kim, *J. Mol. Biol.* **145**, 607 (1981); J. M. Rosenberg, R. E. Dickerson, P. J. Greene, H. W. Boyer, *ibid.* **122**, 241 (1978).
47. R. T. Sauer, R. R. Yocum, R. F. Doolittle, M. Lewis, C. O. Pabo, *Nature (London)* **298**, 447 (1982).
48. K. E. Drexler, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5275 (1981); K. M. Ulmer, in *Molecular Electronic Devices*, F. L. Carter, Ed. (Dekker, New York, 1982), p. 213.

Prospects in Plant Genetic Engineering

Kenneth A. Barton and Winston J. Brill

Agriculture is both the oldest and the largest of the world's industries. Over a period of thousands of years, a broad spectrum of interacting natural and artificial selective pressures has influenced the evolution of crop plants toward those now found under cultivation. Throughout this evolutionary period efforts have been directed toward increasing crop quality and productivity without understanding the contributing molecular features. While the supply of available nutrients for human consumption worldwide has never been in excess, increases in agricultural productivity within the past few decades have been dramatic. A significant reason for the successes of modern agriculture has been an increased reliance on advanced technology, and a successful integration of new technology with the results of intensive plant breeding programs. However, methods of crop improvement in the past few decades have initiated a series of new problems that are now becoming recognized.

Summary. The functional expression of a novel gene in a genetically engineered plant has not yet been reported. One major barrier in movement toward this goal is our limited understanding of the molecular bases of gene expression. Attempts to establish genetic engineering as a practical facet of plant breeding are also complicated by the fact that genes for most important plant characteristics have not yet been identified. However, the benefits to be gained from all aspects of plant improvement are stimulating research into both the development of plant transformation technology and the isolation and characterization of genes responsible for valuable traits. As scientists develop greater knowledge of plant molecular genetics, we can expect to see practical applications in such diverse areas as improvement of plant nutritional quality, decreases in fertilization requirements, and increases in resistance to environmental stresses and pathogens.

Genetically superior plants derived from modern crop improvement programs typically require a high level of crop management. Included in a management regime may be the input of increasingly expensive nitrogen fertilizer as well as the extensive use of pesticides and herbicides, all of which can result in toxic residue accumulation in the environment. In addition, the high degree of inbreeding and the narrowing of the genetic base of widely cultivated crops cause increasing concern about the susceptibility of crops to major disease outbreaks and imply that important genetic traits may be lost as world germplasm is reduced (*1*). With problems such as these it is not surprising that the advent of recombinant DNA technology is generating excitement. A whole range of very specific plant genetic modifications can now be considered, with the use of methods that may someday generate a genetic diversity not naturally present in cultivated plants.

The molecular genetics of prokaryotic organisms is extremely complex and in many respects poorly understood. The flow of stored genetic information in nucleic acids to the appearance of functional gene products elsewhere in the cell requires completion of an intricate sequence of events, with many points where positive or negative control over expression can be exerted. Genetic regulation present in simple eukaryotes, such as yeast, can be more complex, with the added potential for various interactions between organelles, and with an increasing number of both nuclear and cytoplasmic genes. Higher eukaryotes, among them crop plants, provide the still greater problems of cellular differentiation; for example, thousands of active and interacting genes in a leaf cell may be totally quiescent in a root cell of the same organism (*2*). The same natural laws that govern the expression of DNA placed in new genetic environments through classical plant breeding apply to the expression, or lack of expression, of DNA placed in plants by recombinant DNA technology. To be successful in plant genetic engineering, we must begin to develop an understanding of the elements that control gene expression. The significance to gene expression of precise DNA constructs is now beginning to be understood in bacterial, yeast, and

Dr. Barton is a molecular geneticist and Dr. Brill is director of research at Cetus Madison Corporation, Middleton, Wisconsin 53562.