

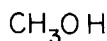
Structure and Function in Biochemistry

by John H. Richards

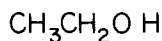
STRUCTURE DETERMINES FUNCTION.

From the simplest arrangements of a few atoms in a small molecule, such as methyl or ethyl alcohol, to molecules with the exquisite and varied architectures of proteins, the molecular structures mandate uniquely and unambiguously their incredibly diverse properties (as drugs, antibiotics, biological catalysts, hormones, transport agents, cell surface receptors, structural elements such as bones or cartilage, muscles that convert chemical energy into work).

Two examples will show how widely different functions characterize molecules with closely related structures. Methyl alcohol (a toxic agent that can cause blindness) and ethyl alcohol (a mild narcotic that can enliven many a party) differ from one another only by one carbon and two hydrogen atoms. In more complex mole-

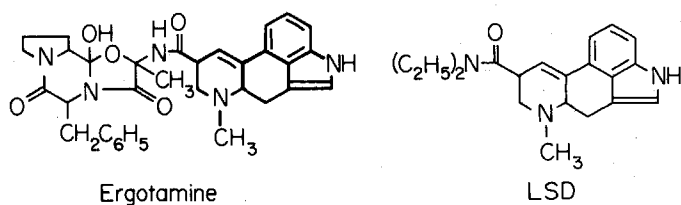


Methyl Alcohol



Ethyl Alcohol

cules, ergotamine finds medicinal utility as a vasoconstrictive agent in treating migraine headaches, whereas LSD, which in structural terms represents about one-half the ergotamine skeleton, has protean effects as a mind-bending agent.



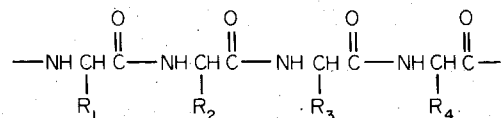
The molecular structure of LSD is identical to that of those regions of ergotamine emphasized in heavy lines.

Over the last few decades, organic chemists have been wonderfully successful in developing an extensive array of pharmaceutical agents without which modern medicine as we know it would not exist: antibiotics for infectious diseases, anti-inflammatory agents for arthritis, drugs to lower high blood pressure and alleviate congestive heart failure, diuretics to accelerate kidney function, anesthetics, tranquilizers and mood elevators, birth-control pills, and chemotherapeutic agents to combat cancer (some of which have turned the

common form of childhood leukemia from a universally fatal disease into one that now has an apparent cure rate of over 60 percent). The possible list is almost endless. In some cases these life-saving drugs have been isolated as natural products from organisms in the world around us — from plants, fungi, bacteria, animals; in many more cases the synthetic art of the organic chemist has proved an essential aspect of the creation of these life-saving or life-enhancing molecules. The indispensable feature that has made all this possible is the present-day ability of the chemist to synthesize almost any organic molecule whose structure holds promise of producing some desirable effect in treating disease.

Until very recently, however, we have lacked the ability to produce specific variations of the molecules that lie at the very heart of every living organism: genes and proteins.

Genes are linear sequences of nucleotides of which there are four types — A (adenine), T (thymine), G (guanine), and C (cytosine). They contain the information that directs the development and functioning of every organism. Information in genes becomes action when translated into proteins. These complex molecules have elaborate and exquisite three-dimensional structures that are determined by the linear sequence of amino acids, of which there are 20 different kinds. This sequence of amino acids, in turn, is determined by the sequences of bases in the genes. Thus one-dimensional information in the form of linear base sequences in the DNA of genes is translated into linear sequences of amino acids in proteins. The basic laws of chemistry and physics then operate on this sequence of amino acids to cause the protein to fold into a precise and beautifully complex three-dimensional structure that posses-



A section of protein. The character of a particular protein is defined by the nature of the side chains (R_1 , R_2 , R_3 , R_4) of each amino acid residue. For this reason subsequent descriptions of changes in protein structures focus on differences in side chains.

ses a very particular biochemical function. In this way the one-dimensional information of genes is converted into functional, three-dimensional proteins.

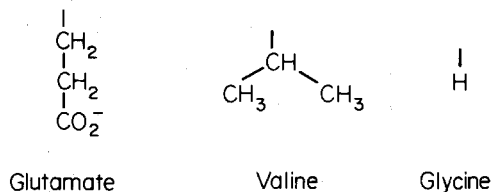
For the reasons just outlined, the secret of generating many structural variants of a protein is to be able to alter, in a precisely controlled manner, its sequence of amino acids — and thereby possibly alter its three-dimensional structure and its functional properties. In this way one can approach in a truly rational manner the way the structure of a protein determines its function. By analogy with the contributions of organic chemists in synthesizing many thousands of molecular structures in the spectacularly successful quest for useful agents in medicine and agriculture, we can anticipate that an ability to generate, at will, proteins with any possibly interesting structure will yield an equally rich harvest of useful molecules.

How then can we create proteins with prespecified sequences of amino acids? Today we have procedures that allow us to achieve this objective. This success derives from recent advances in two areas: molecular biology and synthetic organic chemistry. Modern molecular biology has taught us how to introduce essentially any piece of DNA into a microorganism and cause therein the synthesis of the protein that its nucleotide sequence encodes. This is the fundamental advance that has led recently to the production in abundant quantities of such hitherto exceedingly rare but powerful and useful proteins as human insulin, growth hormone, and interferon. Modern synthetic organic chemistry has enabled us to synthesize, rapidly and easily, sequences of nucleotides that constitute pieces of genes. (Indeed, we can synthesize an entire gene, but this represents still a major, very time-consuming endeavor.) The pieces of genes thus synthesized can then be used to alter, in a way precisely specified by the synthetic gene fragment or oligonucleotide, the sequence of bases in the gene for the parent protein — and thereby generate a modified protein with a sequence of amino acids, and therefore a structure and function, never before available.

This ability is tantamount to creating very specific mutants of normal proteins and is formally termed *oligonucleotide directed mutagenesis*. Not only does this lead to proteins with any structure we may specify, but once a single molecule of the gene encoding that protein has been prepared, the protein itself can be produced forever after in microorganisms by the techniques of modern genetic engineering in whatever quantities may be desired.

In nature, mutations occur spontaneously and provide the variations that, together with selective

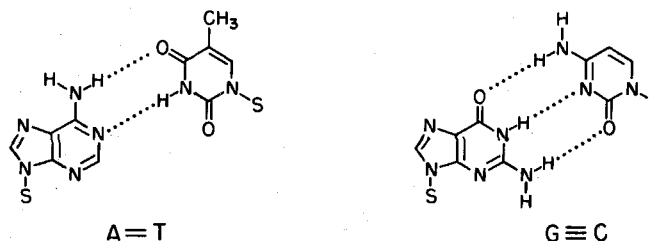
pressures, drive evolution. Even if the mutation is the change in only a single base in a gene with a consequent change of a single amino acid in the protein encoded by that gene, such spontaneous mutations can have profound effects. In the case of the abnormal hemoglobin synthesized by patients with sickle cell anemia, the change is from an A in the codon for glutamic acid to a T in the codon for valine. The mutation of glutamic acid to valine causes the abnormal hemoglobin within red blood cells to gel when it delivers oxygen to the tissues; this gelation in turn causes the deoxygenated cells to sickle, producing the excruciatingly painful crises that characterize this disease. In another example, the change of a single glycine residue to valine in one of the many proteins synthesized by a line of cells grown in tissue culture, causes the cells to become transformed; no longer dividing in a normal controlled manner, they have acquired the characteristics of cancer.



The specific mutations created by oligonucleotide directed mutagenesis involve a combination of organic and biochemical synthesis. Organic synthesis leads to the mutagenic oligonucleotide — generally around 15 nucleotides long — and has become so easy that the task can now be performed by a machine in about 8-10 hours.

The biochemical synthesis is more complex. In the first place, it depends on the necessity for a double-stranded region of DNA as a site at which appropriate enzymes can initiate the replication of DNA, producing thereby a new strand that is the complement of the parent strand. Complementary bases are those pairs that form strongly attractive hydrogen bonds between each other; the recognition of this complementarity of adenine with thy-

Base Pairing

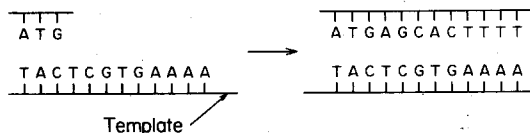


Glutamate is present at the sixth position in the β-chain of normal adult hemoglobin; valine is present at this position in sickle cell hemoglobin. This is the only change between the two hemoglobins and is the ultimate molecular cause of the ravages of sickle cell anemia. The change from glycine to valine in one of the many proteins produced in a line of cells causes the cell line to become cancerous.

The specificity of Watson-Crick base pairing between A = T and G = C accounts for the preservation of the information content of a gene when it is duplicated.

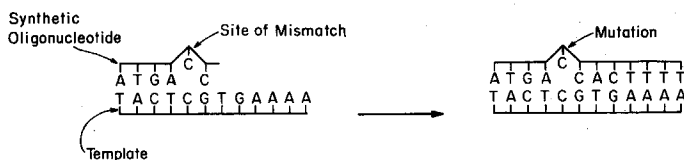
mine and guanine with cytosine played a major role in revealing to Watson and Crick the structure of DNA as a double helix. Thus if one exposes a synthetic oligonucleotide to a long stretch of template DNA that might, for example, constitute a gene, the synthetic oligonucleotide will search along the single-stranded template DNA until it finds a region where its bases complement those of the template, and there the synthetic oligonucleotide will hybridize to form a double-stranded region. The specificity of the complementarity and the rarity of a particular sequence

The biosynthesis of a new strand of DNA uses the information of the template to direct the incorporation of particular nucleotides according to the rules of Watson-Crick base pairing (A-T and G-C).




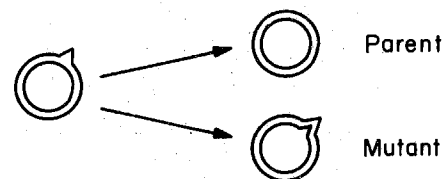
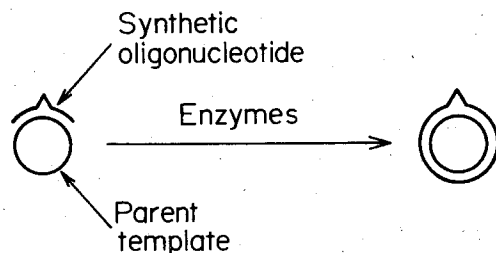
of bases is so high that a synthetic oligonucleotide 15 bases long will find only a single site on a gene many thousands of bases long where all, or most all, of its bases can form Watson-Crick A = T and G = C base pairs. The double-stranded region that results from this hybridization will now serve as an initiation site for enzymes that can complete the synthesis of a strand of new DNA that is complementary to the template.

The secret of introducing a mutation with the synthetic oligonucleotide is that, while most of its bases are indeed complementary to those on the template strand at the site where the oligonucleotide hybridizes, some are not. And these will be incorporated into the newly synthesized strand,



which will accordingly be an exact complement of the parent everywhere except at those sites where the synthetic oligonucleotide had, by design, non-complementary bases. There results consequently a double-stranded DNA, one of whose strands, the parent template, contains the message for the parent protein; the other strand

After completion of the new strand (as above), a heteroduplex plasmid will have been formed, one strand of which contains the information for the parental protein; the other strand carries the information for the new, mutant protein. The site of the mutation is indicated as .



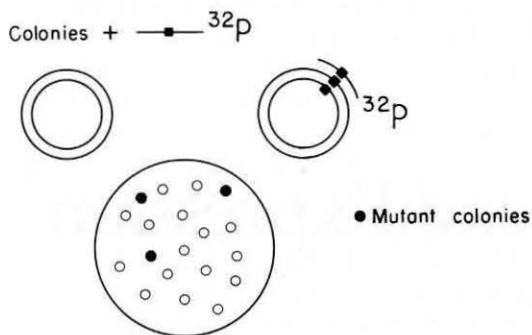
Normal semiconservative replication of the heteroduplex plasmid synthesized in vitro as described leads to two types of descendants, parental and mutant.

contains the message for the mutant gene whose DNA sequence reflects that of the synthetic oligonucleotide. These mutations can be conveniently created in structural genes that are constituents of circular, double-stranded, extrachromosomal plasmid DNA. Plasmids duplicate, as do chromosomes, before cell division; each daughter cell contains one copy of the chromosome of the parent and several copies of the plasmid. The information in the structural genes of the plasmid is translated into protein so that the cell synthesizes proteins encoded both by its chromosomal and plasmid DNA.

After this heteroduplex plasmid DNA is introduced into a suitable bacterial host, such as *E. coli*, the normal semiconservative duplication of double-stranded DNA will lead to two homoduplex descendants. One of these homoduplexes will be like that of the parent template; it will encode the original protein. The other homoduplex, however, will contain the base changes mandated by the mutagenizing synthetic oligonucleotide; it will encode a protein with altered amino acids at those sites specified by the base changes. Thus some clones of the bacterial hosts will contain the message for the parent protein, and some will contain the message for the mutant protein.

We now need to know which are which and could in principle decide by isolating proteins from many such clones and determining which produce mutant and which produce parent protein. But this would be an exceedingly inefficient approach. A much more elegant screening method is available that essentially allows us to "read" the DNA sequence of many clones very rapidly. Recall that the synthetic oligonucleotide will hybridize to the DNA of parental clones with one or more base mismatches (at the sites where the synthetic oligonucleotide purposely contained non-parental bases). In contrast the synthetic oligonucleotide will hybridize perfectly, with no base mismatches, to the DNA of the mutant clones. In hybridizing the mutagenizing oligonucleotide to the parental template to achieve mutagenesis, conditions of relatively low stringency were chosen such that extensive hybridiza-

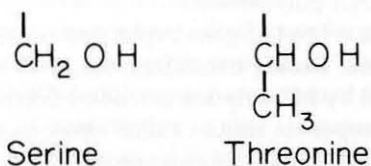
Screening for Mutants by Genotype



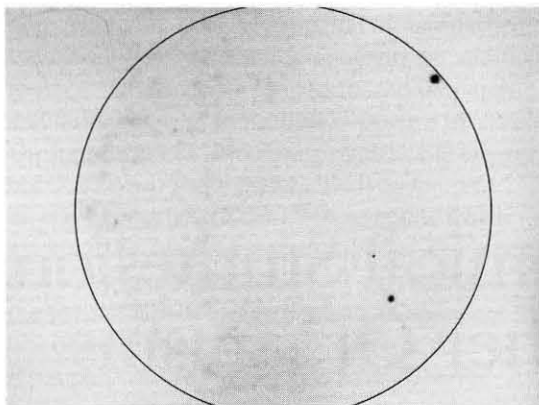
tion occurred even with 1 or 2 base mismatches out of a total of 15 base pairings (that is, 13-14 Watson-Crick base pairs and 2 or 1 mismatched pairs). For screening purposes, much more stringent conditions for hybridization are chosen that allow rapid and unambiguous identification of those clones that contain mutant DNA as compared to those clones that contain parental DNA, *even though the two differ in sequence by as little as a single base in many thousand*. Thus the mutant colonies can be identified, and from them the mutant protein can be isolated.

A recent spectacular application by Dr. Bruce Wallace and his group at the City of Hope of this ability to use the hybridization of a relatively short oligonucleotide to read the DNA sequence of genetic material has allowed the identification, by examination of their DNA, of those individuals who have either one or two genes that encode sickle cell, as contrasted to normal adult, hemoglobin. The genes differ from each other only in a single base. Such screening can be carried out on DNA obtained by amniocentesis and allows the determination early in pregnancy of a genetic defect as small as a single base alteration in the 6×10^9 base pairs that constitute the human genome.

The procedure to produce specifically mutated proteins has been developed at Caltech in collaboration with Dr. Arthur Riggs at the City of Hope and has been applied to the generation of mutants of β -lactamase, an enzyme that efficiently catalyzes the hydrolytic inactivation of penicillin and that is thereby responsible for the resistance to penicillin therapy of an ever increasing number of strains of infectious microorganisms.



The change of a serine to a threonine residue at the active site of β -lactamase leads to a complete loss of catalytic activity.



Under stringent conditions the synthetic oligonucleotide, labeled with radioactive ^{32}P , binds preferentially to those colonies containing mutant plasmid. When the plate is exposed to film, mutant colonies, having bound the radioactive oligonucleotide, appear as dark spots. The photograph represents a real experiment in which two mutant colonies were identified by this screening technique against a background of about 2000 colonies containing plasmid for parental β -lactamase.

In these mutants we altered serine residues to threonine residues in the active site of the enzyme. Serine and threonine differ from each other by only one carbon and two hydrogen atoms (as do methyl and ethyl alcohol). The normal enzyme has a ser-thr dyad at the active site. One mutant had thr-thr, that is, the serine was enlarged by a CH_2 group — a small increment of 14 daltons to a protein of 29,400 daltons. (A dalton is the unit used to express atomic mass.) This mutant is totally devoid of catalytic activity. Another mutant has ser-ser, that is, the threonine was reduced in size by a CH_2 group. This mutant has about 5-10 percent of the catalytic activity of the parent.

These are examples of the very few mutant proteins that have so far been produced by the technique outlined in this article, a technique that has been used to study the relation between protein structure and function for just about one year. So this approach is still in its very early days, and one can anticipate that the future will hold many fascinating surprises. Applications of this technique allow for the first time a truly systematic approach to the central question of the relation between structure and function in proteins and allow us to address such intriguing questions as: What are the essential structural features for a protein to function effectively as a catalyst, hormone, or receptor? How are the various structural aspects of an antibody related to its ability to distinguish self from non-self and thereby, for example, to initiate the destruction of infectious agents such as bacteria and viruses, or the altered cells that form tumors? What regions of a protein located on the surface of a cell carry the information essential to the interactions between cells that play a central role in the development from a single cell (the fertilized egg) of a complex organism such as man? From these studies will come striking new insights into the mechanisms by which proteins successfully carry out their myriad functions as well as the ability to design proteins with specific, novel, and useful properties. □