Protein Structure: Systematic Alteration of Amino Acid Sequences

Efforts to understand how a protein's structure and function depend on its amino acid sequence have yielded little because there has been no systematic way to obtain particular variants of a protein. Investigators may often suspect that a specific amino acid is vital to the structure of a particular protein. But to test that suspicion they had to wait until they, by chance, found mutant cells that produced a variant of the protein in which the amino acid in question, and it alone, was replaced by a different amino acid. Such mutant cells have been difficult, if not impossible, to find.

Recently, Jeffrey Miller and his associates at the University of Geneva and Ponzy Lu of the University of Pennsylvania announced the discovery of a systematic way to isolate specific variants of a protein. In the past few years they have isolated more than 300 variants of the lac repressor protein of the bacterium *Escherichia coli*. In contrast, investigators in a large number of laboratories have found only about 200 variants of human hemoglobin, which is about the same size as the lac repressor protein, although they have looked for such variants for more than two decades. Miller and Lu attribute the rapidity with which they obtained their results to the fact that their systematic technique enables them to insert known amino acids to replace naturally occurring amino acids in the lac repressor protein. They are now using their variants to determine which amino acids are necessary for which functions of the lac repressor and how changes in those amino acids affect the lac repressor's structure and function.

The method of Miller, Lu, and their associates depends in large part on the effects of nonsense mutations, which were discovered more than 10 years ago but never before used in this context. Every time a protein is made from a gene carrying a nonsense mutation the protein's synthesis terminates prematurely. This occurs because the nucleotide sequence of the gene is changed by the nonsense mutation so that a group of three nucleotides (a codon) that normally codes for an amino acid codes instead for a stop signal for gene translation during protein synthesis. For example, one of the three nonsense codons is UAG (for uracil, adenine, and guanine). One of the two codons for the amino acid tyrosine is UAU. Thus a mutation in a tyrosine codon that results in the final uracil being changed to a guanine will transform the tyrosine codon into a stop signal for protein synthesis.

Although synthesis of a protein in normal bacterial cells always stops when a nonsense codon is encountered, certain strains of mutant bacteria will sometimes suppress nonsense codons by inserting an amino acid in a protein at the site of the nonsense codon. Each suppressor strain is specific for a particular nonsense codon and each suppressor strain misreads its own particular nonsense codon as a specific amino acid codon a certain percentage of the time. For example, there are five different suppressor strains that misread the nonsense codon UAG. One of these inserts the amino acid serine in a protein's amino acid sequence instead of stopping at the codon UAG. The other UAG suppressor strains insert different amino acids.

How the Variants Were Obtained

Miller, Lu, and their associates use nonsense mutations and suppressor strains to insert known amino acids at specific positions in the lac repressor. First, they used mutagenic agents to generate nonsense mutations at more than 80 different positions in the lac repressor gene (which codes for a protein consisting of 347 amino acids). Then they transferred each gene carrying a nonsense mutation onto an episome, which is a small piece of DNA that can be taken up by a bacterial cell. Each episome carrying a gene with a specific nonsense codon is used to infect bacteria of the suppressor strains that misread that particular nonsense codon. Thus these investigators know, in each experiment, exactly which amino acid will be replaced in the lac repressor protein (because they know which codon is mutated to a nonsense codon) and which amino acid will replace it (because they know which amino acid each suppressor strain inserts at the site of a nonsense codon). One advantage of this method is that different amino acids can be compared at each site studied. Therefore, at each of the 80 different positions in the repressor gene where nonsense mutations have been found, five different amino acids can be inserted.

Miller, Lu, and their associates point out, though, that there are several disadvantages to the exclusive use of nonsense mutations to obtain variants of a protein. For example, not every codon is easily converted to a nonsense codon by a single mutation. In addition, the number of substitutions that can occur at a nonsense site and the particular amino acids that can be substituted there are restricted by the ways suppressor strains misread nonsense codons. Thus these investigators suggest that collections of variants obtained from suppressions of nonsense mutations be supplemented by variants obtained with more traditional methods.

The relation between the structure and function of the lac repressor has long been of interest to molecular biologists. This protein controls expression of the so-called lac genes in *E. coli* and has become the prototype of proteins that behave in this way. The lac repressor consists of four identical subunits that bind together. Then the resulting tetramer either binds to a specific region of the bacterial DNA (the operator), whereupon it prevents the expression of a particular set of genes, or it binds to certain sugar molecules, whereupon the lac repressor protein changes shape and can no longer bind to DNA.

Miller, Lu, and their associates used their variants of the lac repressor to confirm and extend previously reported results of studies of the structure and function of this repressor. They verified, for example, that substitutions of many amino acids at the amino terminus of the lac repressor affect its binding to DNA but not its binding to sugars. On the other hand, several amino acids in the middle of the protein are indispensable for sugar binding. Other amino acid substitutions actually increase the affinity of the repressor for the operator.

Certain amino acids are found to be crucial to the maintenance of the structure, but not all the functions, of the lac repressor. For example, if some amino acids are replaced, the lac repressor proteins will not aggregate into tetramers but will remain as individual units. These monomers, however, will still bind sugar molecules. Other substitutions failed to produce any detectable changes in repressor functions. The precise knowledge of these amino acid changes demonstrates various relations between the sequence and function of the lac repressor. Miller, Lu, and their colleagues expect that their new collection of variants should prove especially valuable to crystallographers who are now studying the structure of the lac repressor protein with x-ray diffraction.

The method of obtaining variants of a protein by suppressing nonsense mutations depends on a detailed knowledge of the sequence of a protein, and also on a highly developed genetic characterization of the respective gene. Such knowledge is not yet available for proteins of mammalian cells. Even if it were available, the ethics of putting a mammalian gene on a bacterial episome or virus has been questioned. Lu points out, however, that such knowledge is available for many proteins of *E. coli* and for some proteins of bacterial viruses; hence these proteins are prime candidates for application of the new way to study the relation between protein sequence and structure.—GINA BARI KOLATA