Lac System: New Research on How a Protein Binds to DNA

Specific interactions between proteins and DNA sequences are fundamental to the regulation of the metabolism, replication, and development of cells and viruses. For example, the synthesis of certain bacterial enzymes is controlled by proteins that bind to DNA and block or stimulate the transcription of DNA into messenger RNA. The replication of the bacteriophage T4, a bacterial virus, is dependent on a protein that binds to and separates the two strands of the viral DNA, and the response by hen cells to steroid hormones is initiated when a protein interacts with cellular DNA.

In order to understand how proteins interact with DNA to control gene expressions, investigators are studying several model systems. One such system is the lac system of the bacterium *Escherichia coli*. The lac system includes a sequence of genes (the lac operon) whose expression is prevented when a protein binds to a short sequence of DNA that is adjacent to those genes. The protein (the lac repressor), which consists of four identical subunits, binds to a sequence of 21 nucleotides (the lac operator) out of the 6 million nucleotides of *E. coli* DNA.

The DNA of the lac system also includes a control region that contains sites that are apparently bound by two other proteins. (Binding in vitro has never been directly demonstrated, but there is indirect biochemical and genetic evidence that the proteins bind to these sites.) The two proteins are the RNA polymerase (the enzyme that catalyzes the transcription of DNA into messenger RNA) and CRP (that is, the protein that binds cyclic adenosine monophosphate and which must interact with DNA before transcription of the lac genes can begin).

The sequence of the lac repressor protein was recently published by B. Müller-Hill and his colleagues at the University of Köln in Germany at the same time as Walter Gilbert and his colleagues at Harvard University published the sequence of the lac operator DNA and the sequences of two mutant lac operators that are not bound by the lac repressor protein. William Reznikoff and Wayne Barnes of the University of Wisconsin in Madison, together with John Abelson and Robert Dickson of the University of California in San Diego, have reported that they have determined the tentative sequence of the DNA of the control region of the lac system and have confirmed Gilbert's sequence of the lac operator DNA. Others are studying the structure of the lac repressor protein and the differences between the normal lac repressor and mutant lac repressors that are unable to bind to lac operator DNA. Discussions are now beginning to center on specific models of interaction between proteins and DNA in terms of the lac system.

Two types of models have been proposed to explain interactions between proteins and DNA: either the DNA double helix unwinds and the protein binds to single strands of DNA or the protein binds to double helical DNA by binding in a groove of the helix. Both models and a combination of the two have been adapted to explain binding of lac operator DNA by lac repressor protein.

One model that involves a local unwinding of the DNA is a model that was first proposed by Alfred Gierer of the Max-Planck Institut für Virusforschung in Tübingen, Germany, Gierer suggested that the DNA helix could, in some circumstances, unwind and form single-stranded loops and that proteins could bind to the nucleotides of such loops. Loops would form from sequences of DNA that are adjacent to symmetric sequences (Fig. 1). The symmetric sequences of each DNA strand would pair and the configuration could be stabilized when a protein bound to the loops.

The Gierer model was modified to describe the lac system by Henry Sobell of the University of Rochester. Sobell suggested that, since the lac repressor protein consists of four identical subunits, it may bind to four identical sequences of DNA. He proposed that the lac operator DNA may form a structure consisting of two single-stranded loops with identical double-stranded stems and that those stems would be the same as the adjacent double-stranded DNA above and below the stems. Thus the lac operator DNA would comprise four identical double-stranded regions. Each subunit of the lac repressor protein could bind to such a region.

Sobell's model implies that the lac operator DNA would have fourfold symmetry. However, upon determining the sequence of this DNA, Gilbert and his colleagues Allan Maxam and Nancy Maizels found twofold, but not fourfold, symmetry in the sequence (Fig. 1).

Gilbert's group notes that single-stranded loops could form from the symmetric regions of the lac operator DNA, although the stems of those loops would not be identical to the adjacent double-stranded DNA. The repressor protein could possibly bind to such loops, but Gilbert and his colleagues believe that an evaluation of the importance of symmetric DNA sequences to binding by repressor protein will follow from comparisons between normal lac operator DNA and mutant lac operator DNA, which is not bound by lac repressor protein. This mutant DNA must lack features that are essential to binding by lac repressor protein.

Gilbert and his associates have determined the sequence of the lac operator regions of two mutants in which binding by lac repressor protein is greatly reduced. In each of the mutants, one nucleotide pair of the lac operator sequence is changed. In one mutant, a nucleotide pair of a symmetric region is changed. This could alter the stability of the stem of a loop. The second mutant contains an altered base pair in a nonsymmetric region. As Gilbert points out, this second explanation is easily explained by Gierer's model since an assumption of that model is that the protein would bind to the nucleotides of single-stranded loops.

Additional evidence that symmetric DNA sequences may be important to binding by proteins of the lac system follows from the proposed sequence of the DNA of the control region. Reznikoff, Abelson, and their associates found that the symmetry of the lac operator DNA is extended by the four nucleotides that are adjacent to the lac operator region so that 12 of the 15 base pairs on either side of the center of the operator region are symmetric.

Reznikoff and Abelson have offered evidence of the location of those DNA sequences that are thought to be bound by the RNA polymerase and the CRP. They compared the sequences of DNA from normal cells to that of mutants in which the DNA of that region appears not to be bound by RNA polymerase and mutants in which the DNA of that region appears not to be bound by CRP. The DNA se-

quence that is apparently bound by RNA polymerase seems to include two copies of a sequence of five nucleotides, but it is not a symmetric sequence. The RNA polymerase consists of four subunits. Two of these subunits are identical, but it is not known whether the identical subunits are necessary for binding to this region of DNA. Thus the type of symmetry in the sequences of the DNA that is bound by proteins of the lac system is consistent with models that include looped structures, but in itself does not constitute sufficient evidence to rule out other models of interaction between proteins and DNA.

According to a second class of models of interaction, a protein could recognize DNA sequences from outside the double helix. The protein would form contacts with the nucleotide bases through the large or small groove of the DNA helix. Such a model was adapted to describe interactions of the lac system by Müller-Hill and his colleagues.

Müller-Hill proposed that one end of the lac repressor protein, the amino terminus, is in the form of an alpha helix and that this helix protrudes from the repressor and binds to the large groove of the DNA. The helical protrusion would bind to the DNA (i) by means of electrostatic interactions between the negatively charged DNA and the positively charged groups of the repressor protein and (ii) by means of hydrogen bonds between the amino acids of the protruding helix and specific nucleotides of the operator DNA.

By suggesting such a model for the lac system, Müller-Hill hoped to explain how an amino acid might bind to a specific nucleotide and how the amino terminus of the lac repressor protein might be involved in this binding.

Evidence from independent lines of research established the fact that the amino terminus of the repressor protein is necessary for DNA binding. Müller-Hill's group analyzed mutant repressor proteins that could no longer bind to DNA and found that mutations in the amino terminus prevented binding. Terry Platt, now at Stanford University, together with James Files and Klaus Weber of Harvard University, partially degraded normal repressor molecules with trypsin and chymotrypsin. These enzymes degraded the amino terminus of the repressor, indicating that the amino terminus is exposed to the environment. This partially degraded repressor protein could no longer bind to DNA, even though it still consisted of four subunits and still could bind the sugar isopropyl-β-D-thiogalactoside. Müller-Hill's model of specific interactions between the amino terminus of the repressor and the bases of the operator DNA is consistent with the evidence that the amino terminus is necessary for interaction with operator DNA; but it does not provide a unique explanation of this phenomenon.

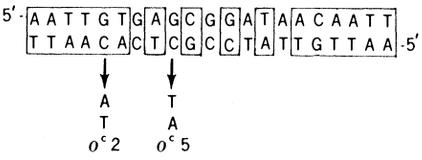

Fig. 1. The sequence of lac operator DNA and the sequences of two mutant lac operator DNA's that are not bound by the lac repressor protein. The symmetric regions in the operator are boxed. The mutation o^c2 is a transition from G · C to A · T. The mutation o^c5 is a transition from G · C to T · A. (G, guanine; C, cytosine; A, adenine; and T, thymine.)

Another way to study the interaction between the lac repressor protein and the lac operator DNA would be to crystallize molecules of the repressor-operator complex and to study the structure of that complex by x-ray diffraction. Although no one has been able to crystallize the repressor-operator complex, Thomas Steitz and his associates at Yale University have obtained small crystals of lac repressor protein and have begun to study its structure.

From their preliminary studies of the structure of the lac repressor protein, Steitz and his colleagues propose a model that describes the interaction between the repressor and the operator DNA. They neither require nor exclude the direct involvement of the amino terminus in DNA binding. They point out that other parts of the repressor may bind to DNA and propose that the repressor recognizes the operator from the outside of the DNA and fits around the DNA like a bun around a hot dog.

Steitz and his associates analyzed the repressor's structure by means of negatively stained electron micrographs and x-ray powder patterns. These studies led them to conclude that each repressor molecule is shaped like a dumbbell or narrow-waisted ellipsoid and to determine that the molecule is 130 to 140 Å long and about 60 by 40 Å in the other two dimensions.

Steitz's group believes that, if the shape of the repressor molecule when interacting with operator DNA does not drastically change, their data rule out Sobell's model as well as Müller-Hill's model of DNA binding. Sobell's model leads to the prediction that the repressor has a molecular fourfold axis, which is inconsistent with the asymmetric shape indicated by the structural studies. Müller-Hill's model requires that the repressor subunits be arranged in a helical fashion around the DNA. If this were the case, Steitz argues, the repressor would have a circular rather than an elliptical profile in the electron micrographs.

Steitz has proposed a model that is consistent with the operator sequence determined by Gilbert's group. He proposes that the repressor binds to DNA with its long axis aligned with the long axis of the DNA and that the DNA binding site should be along the line of intersection of the four repressor subunits and perhaps consist of a groove in the repressor molecule. Thus the DNA binding site would have the twofold symmetry observed by Gilbert, Maizels, and Maxam; and all four repressor subunits could interact with the DNA.

Despite the work on the sequence of the lac operator DNA and the structure of the lac repressor protein and the work on the sequence of the control region of the lac system, the problem of explaining the interaction between the operator and the repressor and between the RNA polymerase and its target site and the CRP and its target site remains unsolved. A resolution of the problem awaits further study of the structure of the lac repressor and the determination of the nucleotide sequences of additional mutants whose lac operator regions are not bound by repressor, further study of mutant lac repressor proteins that are unable to bind to lac operator DNA, and further study of the sequence of the control region. Because such studies are feasible, researchers believe that, for the lac system, a solution of how proteins bind to DNA to control the expression of genes is now within reach.

—GINA BARI KOLATA

Additional Readings

1. K. Adler, K. Beyreuther, E. Fanning, N. Geisler, B. Gronenborn, A. Klemm, B. Müller-Hill, M. Pfahl, A. Schmitz, *Nature (Lond.)* **237**, 322 (1972).
2. W. Gilbert and A. Maxam, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3581 (1973).
3. N. Maizels, *ibid.*, p. 3585.
4. T. Steitz, T. J. Richmond, D. Wise, D. Engleman, *ibid.*, in press.