Perspective

DNA Looping

ROBERT SCHLEIF

DNA STORES THE INFORMATION NECESSARY FOR REPLICATION and subsequent cell growth in the sequence of its constituent nucleotides. As needed, appropriate blocks of this information are read out into RNA copies by RNA polymerase. The blocks of information are called genes, and the initial region to which RNA polymerase binds is called a promoter. Growth requires that the readout be regulated according to the cell's needs. In addition to RNA polymerase, which transcribes the gene, other proteins often regulate the transcriptional readout from genes. These other proteins usually bind to DNA near a promoter. We would expect and, indeed, it has been found that both general and gene-specific proteins bind to DNA in the vicinity of promoters.

Over the years, highly detailed studies of a handful of prokaryotic promoters and their associated regulatory sequences in the bacterium *Escherichia coli* and in bacteriophage lambda revealed that regulatory sequences are immediately adjacent to promoters. Consequently, we have come to view regulation processes as being mediated by one or more proteins that can sense intracellular conditions. These proteins bind DNA beside the gene they regulate, directly interact with one another, and interact with RNA polymerase to stimulate or inhibit transcription of the associated gene. A fundamental question concerning eukaryotic cells is whether they regulate gene activity in the same general fashion. Early studies in eukaryotic cells did identify promoters. Soon however, it became clear that DNA sequences important to the transcription of eukaryotic genes often lie hundreds or thousands of nucleotides away from the genes. These sequences are called enhancers. Their presence raises the perplexing problem of understanding how information can be communicated from an enhancer to its promoter. In the past few years, new work in several prokaryotic systems has provided a plausible mechanism for understanding the action-at-a-distance phenomenon seen with many eukaryotic enhancers.

Well after the discovery of enhancers in eukaryotes, evidence for regulatory sequences located hundreds of nucleotides away from their promoters was also found in *E. coli*. Such sequences were reported for the *gal* operon (*1*), *ara* operon (*2*), *lac* operon (*3*), one of the *deo* operons (*4*), and others. Thus, bacteria also have distantly located regulatory sites. The relative ease of genetic manipulations in *E. coli* facilitated investigation of these sites. Evidence was quickly generated in the *ara* operon that it is DNA looping that provides the communication between the two separated regulatory sites (*2*). Proteins binding to each of the sites also bind to each other at the same time. This generates a loop of DNA held together by the proteins.

What is the evidence for looping? The first evidence obtained was somewhat indirect and is described below. The most direct evidence is electron microscope visualization of loops formed in vitro by two regulatory proteins linking two DNA sites. Such loops were first seen in an artificial looping system constructed with lambda phage repressor and operators (*5*). They have also been observed when special tight-binding operators for *lac* repressor were placed several hundred nucleotides apart (*6*), and in a eukaryotic system consisting of a steroid hormone-receptor and its DNA binding sites (*7*).

Two types of in vivo experiments also demonstrate DNA looping. One of these is the helical twist experiments that were done in the *ara* system and that first illustrated looping (*2*). These experiments are based on the fact that DNA is a three-dimensional cylinder and not the line often depicted in textbooks. Therefore if two proteins are bound to two sites on the same side of the DNA, they can form a loop by bending the DNA so that the proteins contact one another (*8*). If, however, the proteins bind on opposite sides of the DNA helix, forming a loop requires both twisting the DNA between the sites by half a turn so that the proteins are on the same side as well as bending the DNA into a loop. For sites separated by several hundred base pairs or less, such twisting requires 4 kilocalories per mole or more of energy per mole of reactant. This is substantial compared to the 10 to 15 kilocalories per mole of binding energy required for site-specific protein-DNA interactions. Displacement of the two *ara* sites by half-integral numbers of helical turns so that they are on opposite sides of the DNA interferes with looping. Systems with a greater separation between sites or with greater energies available from the combination of the protein-DNA binding interaction and the protein-protein interaction should show weaker helical twist effects.

The second demonstration of in vivo looping is binding site cooperativity. Two sites involved in looping in the *ara* system are $araO_2$ and *araI*. While the binding of the dimeric *ara* regulatory protein AraC to *araI* is tight, binding to $araO_2$ is weak. Nonetheless, measurement of the binding of AraC protein to these two sites by in vivo footprinting shows that normally both are occupied (*9*). However, if *araI* and an adjacent AraC binding site called $araO_1$ are deleted, then $araO_2$ is no longer occupied by AraC protein. That is, the binding of AraC to *araI* assists the binding of another AraC dimer at $araO_2$. Such binding site cooperativity can easily be understood as a consequence of DNA looping. A dimer of the protein bound at *araI* and its potential for binding another dimer to itself substantially increases the effective concentration of a dimer in the vicinity of $araO_2$, a phenomenon known as the chelate effect (*10*). The reaction rate between two molecules in solution can be enormously increased by fastening them together or by holding them in the correct angular orientation with respect to one another. Several factors make calculation of the cooperativity in looping systems difficult, but we estimate that the effect of looping brings the effective concentration of AraC in the vicinity of $araO_2$ to more than $10^{-7}M$ whereas without looping the concentration would be less than $10^{-9}M$.

An elegant set of experiments on in vitro and in vivo looping has been described by Müller-Hill and his colleagues (*3, 6*). For the in vitro experiments the gel retardation assay was used. The electrophoresis rate of a 100– to 1000–base pair (bp) DNA fragment through acrylamide or agarose gels is retarded if a protein is bound to the DNA. Electrophoresis conditions can be found in which many DNA-binding proteins remain bound to their natural DNA binding sites for the 1- to 3-hour duration of an electrophoresis separation. Consequently, if protein binds to a DNA fragment, an easily detected slower moving band forms in the gel. The extraordinary specificity of this assay often permits detection with crude cell extracts of DNA binding proteins.

The gel retardation assay is a sensitive detector of looping in the *lac* system. When two variant *lac* operators with high affinity for *lac*

The author is a professor in the Graduate Department of Biochemistry at Brandeis University, Waltham, Massachusetts 02254.

repressor are separated by several hundred nucleotides and are on opposite sides of the DNA, the expected bands form in the gel. These are DNA molecules with zero, one, and two *lac* tetramers of *lac* repressor bound. When the two operators are on the same side of the DNA cylinder, repressor binding is highly cooperative, and virtually none of the species with a single tetramer is found. Instead, nearly all the DNA is found in a new band of abnormally low migration rate, DNA looped by a single repressor tetramer. At high concentrations of DNA and repressor, another species was found; this was two DNA molecules cross-joined by two repressor tetramers.

An in vivo set of experiments provides an explanation for a long-standing enigma of the *lac* operon. Early experiments on *lac* repressor binding to DNA revealed two pseudo-operator sites in the *lac* operon, each within a couple hundred nucleotides of the operator that overlaps the promoter (*11*). One lies at the end of the *lacI* gene and the other near the beginning of the *lacZ* gene. Are these vestiges of evolution, the results of recombination events, or do they help regulate the operon? The binding site cooperativity generated by DNA looping could increase binding of repressor to operator and more efficiently repress the operon. One clue that they might play such a role has come from cloning experiments in which the *lac* promoter without these pseudo-operators has been used to drive synthesis of other genes. These artificial constructs do not fully repress the *lac* promoter. Likely then, the pseudo-operators in the *lac* system facilitate repression by DNA looping. This was proved when the pseudo-operator in the *lacZ* gene was eliminated by substituting alternative but synonymous codons (*3*). Normal LacZ protein is synthesized, but repression is now inefficient, paralleling that found in the artificial constructs.

I suggest two reasons for the existence of DNA looping. First, looping provides a convenient mechanism for achieving high affinity binding via the binding site cooperativity inherent in the systems. Second, looping facilitates regulating gene activity when more than several regulatory proteins are involved. One or two proteins can bind to DNA alongside RNA polymerase, but having RNA polymerase respond to still more regulatory proteins generates complications. Looping permits additional regulatory proteins to be placed a substantial distance away from the RNA polymerase and still influence the initiation process either by assisting or hindering the loop formation. Multiple looping species are also possible, and have now been observed in the *ara* system (*12*). Apparently regulation mechanisms can also utilize mutually exclusive alternative looping schemes. In summary, DNA looping provides a conceptually simple explanation for the action-at-a-distance phenomenon that is found in prokaryotic and eukaryotic cells. It is a logical and versatile mechanism for regulating gene activity.

REFERENCES AND NOTES

1. M. Irani, L. Orosz, S. Adhya, *Cell* **73**, 783 (1983).
2. T. Dunn, S. Hahn, S. Ogden, R. Schleif, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5017 (1984).
3. E. Eismann, B. Wilcken-Bergmann, B. Müller-Hill, *J. Mol. Biol.* **195**, 949 (1987).
4. G. Dandanell and K. Hammer, *EMBO J.* **4**, 333 (1985).
5. J. Griffith, A. Hochschild, M. Ptashne, *Nature (London)* **322**, 750 (1986).
6. H. Krämer *et al.*, *EMBO J.* 6, 1481 (1987).
7. B. Théveny *et al.*, *Nature (London)* **329**, 79 (1987).
8. An extreme case of two proteins binding to each other is one protein with two DNA binding sites.
9. K. Martin, L. Huo, R. Schleif, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3654 (1986).
10. M. Page and W. Jencks *ibid.* **68**, 1678 (1971).
11. W. Reznikoff, R. Winter, C. Hurley, *ibid.* **71**, 2314 (1974); W. Gilbert, J. Majors, A. Maxam, *Life Sci. Res. Rep.* **4**, 167 (1976).
12. L. Huo, K. Martin, R. Schleif, in preparation.