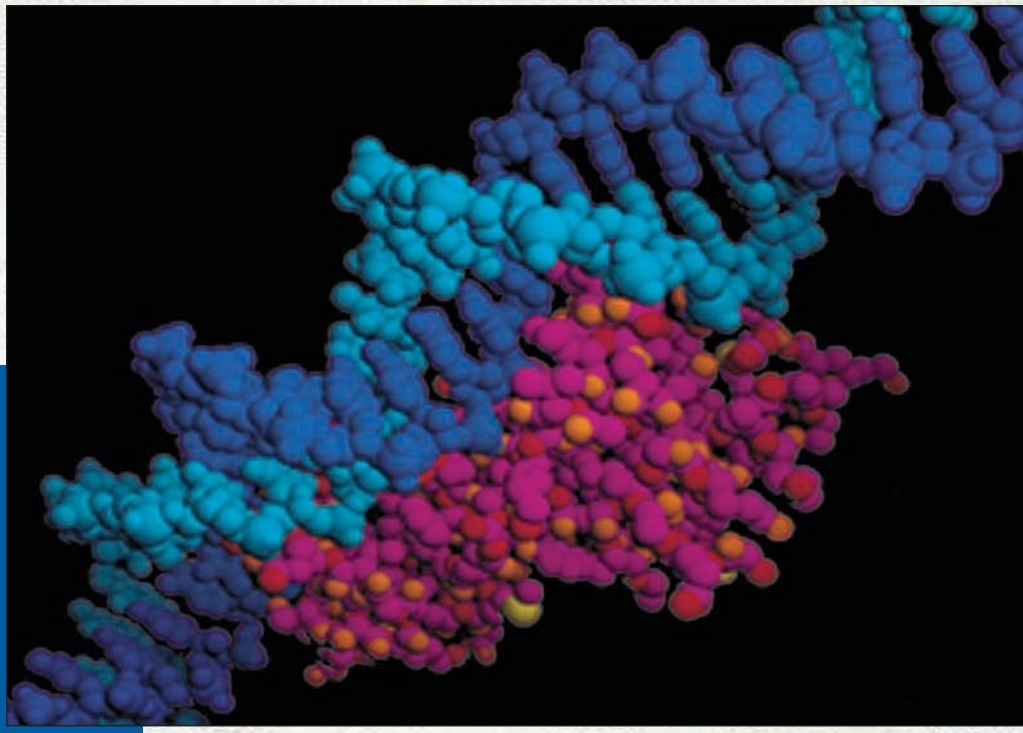
CHAPTER 9

DNA–Protein Interactions in Bacteria



Gene regulation: Computer model of Cro protein bound to DNA.
© Ken Eward/SS/Photo Researchers, Inc.

In Chapters 7 and 8 we discussed several proteins that bind tightly to specific sites on DNA. These include RNA polymerase, *lac* repressor, CAP, *trp* repressor, λ repressor, and Cro. All of these have been studied in detail, and all can locate and bind to one particular short DNA sequence among a vast excess of unrelated sequences. How do these proteins accomplish such specific binding—akin to finding a needle in a haystack? The latter five proteins have a similar structural motif: two α-helices connected by a short protein "turn." This **helix-turn-helix motif** (Figure 9.1a) allows the second helix (the **recognition helix**) to fit snugly into the major groove of the target DNA site (Figure 9.1b). We will see that the configuration of this fit varies considerably from one protein to another, but all the proteins fit their DNA binding sites like a key in a lock. In this chapter we will explore several

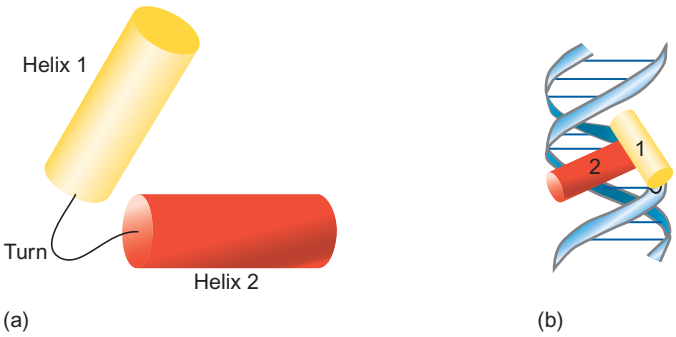


Figure 9.1 The helix-turn-helix motif as a DNA-binding element. **(a)** The helix-turn-helix motif of the λ repressor. **(b)** The fit of the helix-turn-helix motif of one repressor monomer with the λ operator. Helix 2 of the motif (red) lies in the major groove of its DNA target; some of the amino acids on the back of this helix (away from the viewer) are available to make contacts with the DNA.

well-studied examples of specific DNA–protein interactions that occur in prokaryotic cells to see what makes them so specific. In Chapter 12 we will consider several other DNA-binding motifs that occur in eukaryotes.

9.1 The λ Family of Repressors

The repressors of λ and similar phages have recognition helices that lie in the major groove of the appropriate operator as shown in Figure 9.2. The specificity of this binding depends on certain amino acids in the recognition helices

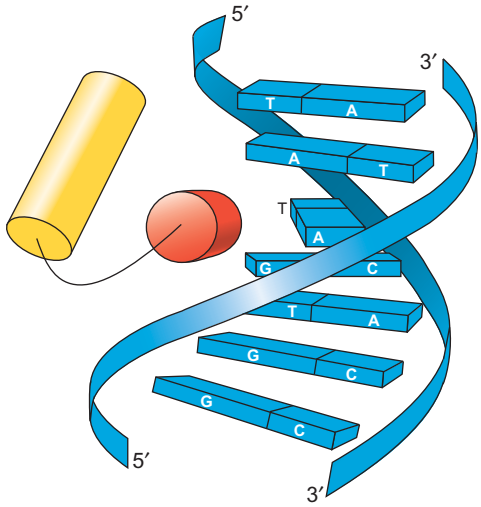


Figure 9.2 Schematic representation of the fit between the recognition helix of a λ repressor monomer and the major groove of the operator region of the DNA. The recognition helix is represented by a red cylinder that lies in the major groove in a position to facilitate hydrogen bonding with the edges of base pairs in the DNA. (*Source:* Adapted from Jordan, S.R. and C.O. Pabo. Structure of the lambda complex at 2.5 Å resolution. Details of the repressor–operator interaction. *Science* 242:896, 1988.)

that make specific contact with functional groups of certain bases protruding into the major DNA groove, and with phosphate groups in the DNA backbone. Other proteins with helix-turn-helix motifs are not able to bind as well at that same site because they do not have the correct amino acids in their recognition helices. We would like to know which are the important amino acids in these interactions.

Probing Binding Specificity by Site-Directed Mutagenesis

Mark Ptashne and his colleagues have provided part of the answer to the specificity question, using repressors from two λ-like phages, **434** and **P22**, and their respective operators. These two phages have very similar molecular genetics, but they have different immunity regions: They make different repressors that recognize different operators. Both repressors resemble the λ repressor in that they contain helix-turn-helix motifs. However, because they recognize operators with different base sequences, we would expect them to have different amino acids in their respective recognition helices, especially those amino acids that are strategically located to contact the bases in the DNA major groove.

Using x-ray diffraction analysis (Box 9.1) of operator–repressor complexes, Stephen Harrison and Ptashne identified the face of the recognition helix of the 434 phage repressor that contacts the bases in the major groove of its operator. By analogy, they could make a similar prediction for the P22 repressor. Figure 9.3 schematically illustrates the amino acids in each repressor that are most likely to be involved in operator binding.

If these are really the important amino acids, one ought to be able to change only these amino acids and thereby alter the specificity of the repressor. In particular, one should be able to employ such changes to alter the 434 repressor so that it recognizes the P22 operator instead of its own. This is exactly what Robin Wharton and Ptashne did. They started with a cloned gene for the 434 repressor and, using mutagenesis techniques similar to those described in Chapter 5, systematically altered the codons for five amino acids in the 434 recognition helix to codons for the five corresponding amino acids in the P22 recognition helix.

Next, they expressed the altered gene in bacteria and tested the product for ability to bind to 434 and P22 operators, both in vivo and in vitro. The in vivo assay was to check for immunity. Recall that an *E. coli* cell lysogenized by λ phage is immune to superinfection by λ because the excess λ repressor in the lysogen immediately binds to the superinfecting λ DNA and prevents its expression (Chapter 8). Phages 434 and P22 are λ-like (lambdoid) phages, but they differ in their immunity regions, the control regions that include the repressor genes and the operators. Thus, a 434 lysogen is immune to superinfection by 434, but not by P22. The 434 repressor cannot bind to the

BOX 9.1

X-Ray Crystallography

This book contains many examples of structures of DNA-binding proteins obtained by the method of **x-ray diffraction analysis,** also called **x-ray crystallography.** This box provides an introduction to this very powerful technique.

X-rays are electromagnetic radiation, just like light rays, but with much shorter wavelengths so they are much more energetic. Thus, it is not surprising that the principle of x-ray diffraction analysis is in some ways similar to the principle of light microscopy. Figure B9.1 illustrates this similarity. In light microscopy (Figure B9.1), visible light is scattered by an object; then a lens collects the light rays and focuses them to create an image of the object.

In x-ray diffraction, x-rays are scattered by an object (a crystal). But here we encounter a major problem: No lens is capable of focusing x-rays, so one must use a relatively indirect method to create the image. That method is based on the following considerations: When x-rays interact with an electron cloud around an atom, the x-rays scatter in every direction. However, because x-ray beams interact with multiple atoms, most of the scattered x-rays cancel one another due to their wave nature. But x-rays scattered to certain specific directions are amplified in a phenomenon called *diffraction.* Bragg's law, $2d \sin \theta = \lambda$, describes the relationship between the angle (θ) of diffraction and spacing (d) of scattering planes. As you can see in Figure B9.2, x-ray 2 travels $2 \times d \sin \theta$ longer than x-ray 1. Thus, if the wavelength (λ) of x-ray 2 is equal to $2 d \sin \theta$, the resultant rays from the scattered x-ray 1 and x-ray 2 have the same phase and are therefore amplified. On the other hand,

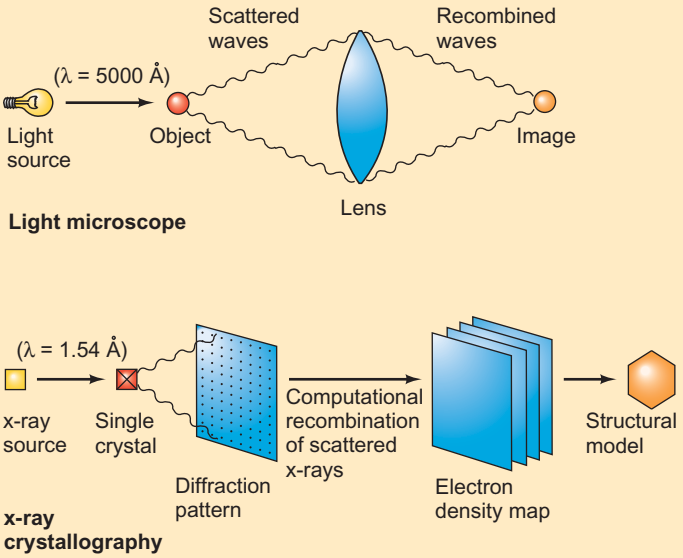


Figure B9.1 Schematic diagram of the procedures followed for image reconstruction in light microscopy (top) and x-ray crystallography (bottom).

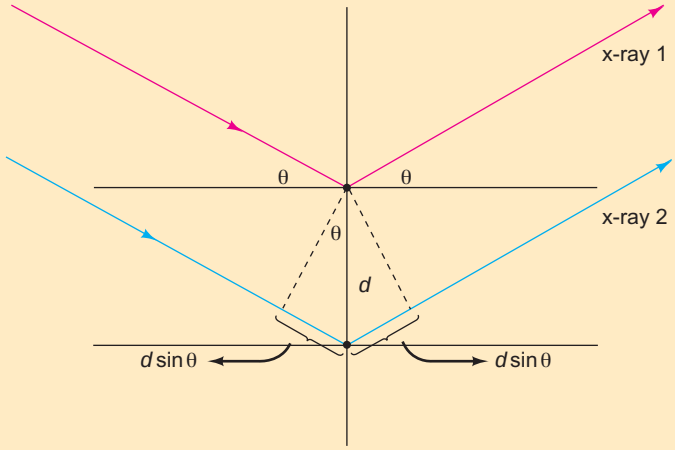


Figure B9.2 Reflection of two x-rays from parallel planes of a crystal. The two x-rays (1 and 2) strike the planes at angle θ and are reflected at the same angle. The planes are separated by distance d. The extra distance traveled by x-ray 2 is $2 d \sin \theta$.

the resultant rays are diminished if λ is not equal to $2 d \sin \theta$. The diffracted x-rays are recorded as spots on a collecting device (a detector) placed in the path of the x-rays. This device can be as simple as a sheet of x-ray film, but nowadays much more efficient electronic detectors are available. Figure B9.3 shows a diffraction pattern of a simple protein, lysozyme. Even though the protein is relatively simple (only 129 amino acids), the pattern of spots is complex. To obtain the protein structure in three dimensions, one must rotate the crystal and record diffraction patterns in many different orientations.

The next task is to use the arrays of spots in the diffraction patterns to figure out the structure of the molecule that caused the diffraction. Unfortunately, one cannot reconstruct the **electron-density map** (electron cloud distribution) from the arrays of spots in the diffraction patterns, because information about the physical parameters, called *phase angles,* of individual reflections are not included in the diffraction pattern. To solve this problem, crystallographers make 3–10 different heavy-atom derivative crystals by soaking heavy atom solutions (Hg, Pt, U, etc.) into protein crystals. These heavy atoms tend to bind to reactive amino acid residues, such as cysteine, histidine, and aspartate, without changing the protein structure.

This procedure is called multiple isomorphous replacement (MIR). The phase angles of individual reflections are determined by comparing the diffraction patterns from the native and heavy-atom derivative crystals. Once the phase angles are obtained, the diffraction pattern is mathematically converted to an electron-density map of the diffracting molecule. Then the electron-density map can be used to infer the

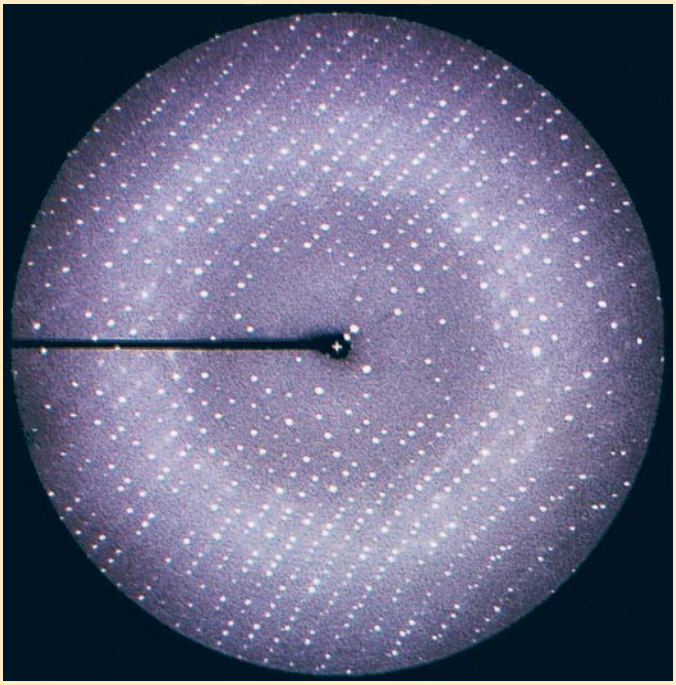
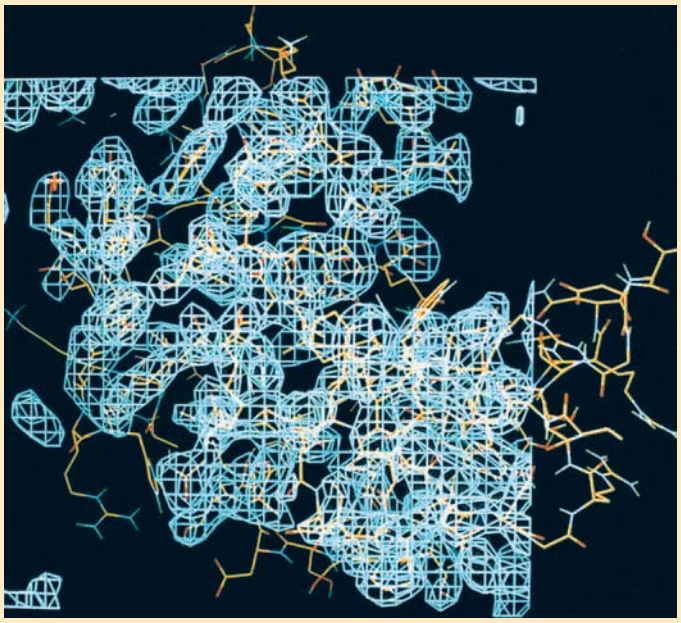


Figure B9.3 Sample diffraction pattern of a crystal of the protein lysozyme. The dark line from the left is the shadow of the arm that holds the beam stop, which protects the detector from the x-ray beam. The location of the crystal is marked by the (+) at the center. (*Source:* Courtesy of Fusao Takusagawa.)

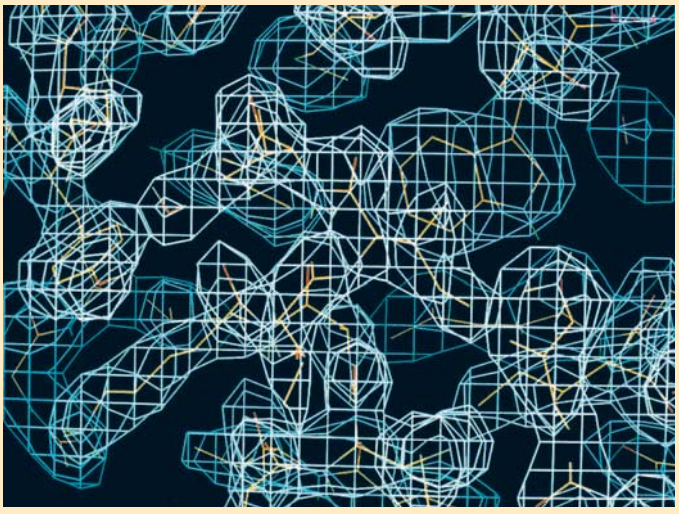
structure of the diffracting molecule. Using the diffracted rays to create an image of the diffracting object is analogous to using a lens. But this is not accomplished physically, as a lens would; it is done mathematically. Figure B9.4 shows the electron-density map of part of the structure of lysozyme, surrounding a stick diagram representing the molecular structure inferred from the map. Figure B9.5 shows three different representations of the whole lysozyme molecule deduced from the electron-density map of the whole molecule.

Why are single crystals used in x-ray diffraction analysis? It is clearly impractical to place a single molecule of a protein in the path of the x-rays; even if it could be done, the diffraction power from a single molecule would be too weak to detect. Therefore, many molecules of protein are placed in the x-ray beam so the signal will be strong enough to detect. Why not just use a protein powder or a solution of protein? The problem with this approach is that the molecules in a powder or solution are randomly oriented, so x-rays diffracted by such a sample would not have an interpretable pattern.

The solution to the problem is to use a crystal of protein. A crystal is composed of many small repeating units (**unit cells**) that are three-dimensionally arranged in a regular



(a)

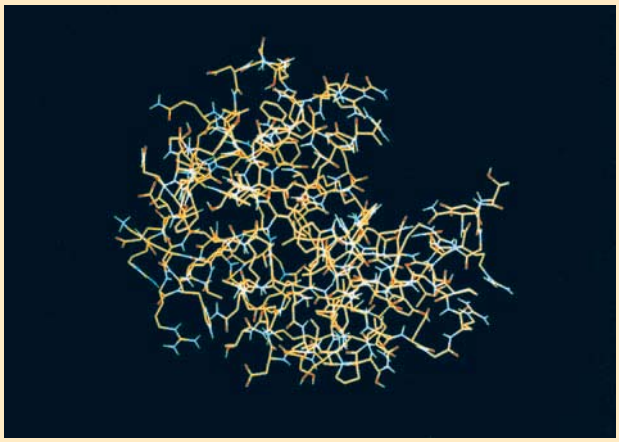


(b)

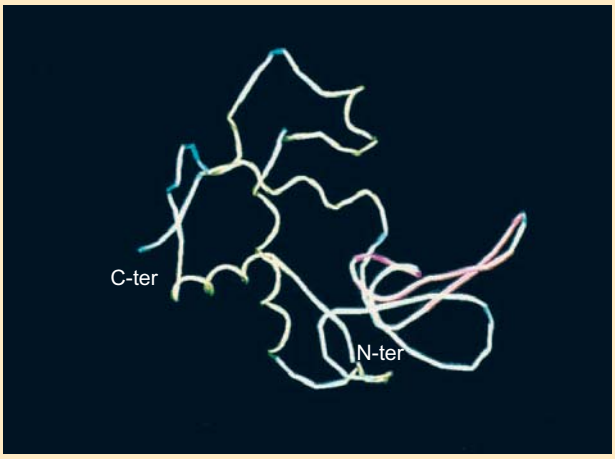
Figure B9.4 Electron-density map of part of the lysozyme molecule. (a) Low magnification, showing the electron density map of most of the molecule. The blue cages correspond to regions of high electron density. They surround a stick model of the molecule (red, yellow, and blue) inferred from the pattern of electron density. **(b)** High magnification, showing the center of the map in panel (a). The resolution of this structure was 2.4 Å so the individual atoms were not resolved. But this resolution is good enough to identify the unique shape of each amino acid. (*Source:* Courtesy Fusao Takusagawa.)

continued

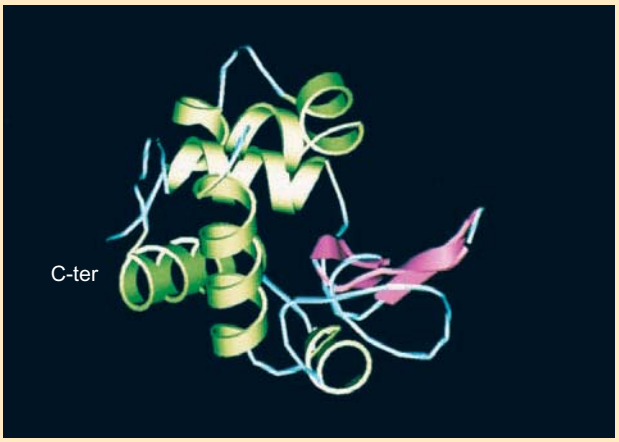
X-Ray Crystallography *(continued)*



(a)



(b)



(c)

Figure B9.5 Three representations of the structure of lysozyme calculated from electron density maps such as those in Figure B9.4. **(a)** Stick diagram as in Figure B9.4a. **(b)** String diagram with α-helices in green, β-sheets in magenta, and random coils in blue. The N-terminus and C-terminus of the protein are marked N-ter and C-ter, respectively. **(c)** Ribbon diagram with same color coding as in panel (b). The helical nature of the α-helices is obvious in this diagram. The cleft at upper right in all three diagrams is the active site of the enzyme. (*Source:* Courtesy Fusao Takusagawa.)

way. A unit cell of a protein contains several protein molecules that are usually related by special symmetries. Thus, diffractions by all the molecules in a unit cell in the crystal are the same, and they reinforce one another. To be useful for x-ray diffraction, the smallest dimension of a protein crystal should be at least 0.1 mm. A cubic crystal of this size contains more than 10^{12} molecules (assuming that one protein molecule occupies a 50 × 50 × 50 Å space). Figure B9.6 presents a photograph of crystals of lysozyme suitable for x-ray diffraction analysis. Protein crystals contain not only pure protein but also a large amount of solvent (30–70% of their weight). Thus, their environment in the crystal resembles that in solution, and their three-dimensional structure in the crystal should therefore be close to their structure in solution. In general, then, we can be confident that the protein structures determined by x-ray crystallography are close to their structures in the cell. In fact, most enzyme crystals retain their enzymatic activities.

Why not just use visible light rays to see the structures of proteins and avoid all the trouble involved with x-rays? The problem with this approach lies in **resolution**—the ability to distinguish separate parts of the molecule. The ultimate goal in analyzing the structure of a molecule is to distinguish each atom, so the exact spatial relationship of all the atoms in the molecule is apparent. But atoms have dimensions on the order of angstroms (1 Å = 10^{-10} m), and the maximum resolving power of radiation is one-third of its wavelength ($0.6\lambda/2 \sin \theta$). So we need radiation with a very short wavelength (measured in angstroms) to resolve the atoms in a

P22 operators and therefore cannot prevent superinfection by the P22 phage. The reverse is also true: A P22 lysogen is immune to superinfection by P22, but not by 434.

Instead of creating lysogens, Wharton and Ptashne transformed *E. coli* cells with a plasmid encoding the recombinant 434 repressor, then asked whether the recombinant 434 repressor (with its recognition helix altered to be like the P22 recognition helix) still had its original binding specificity. If so, cells producing the recombinant repressor should have been immune to 434 infection. On

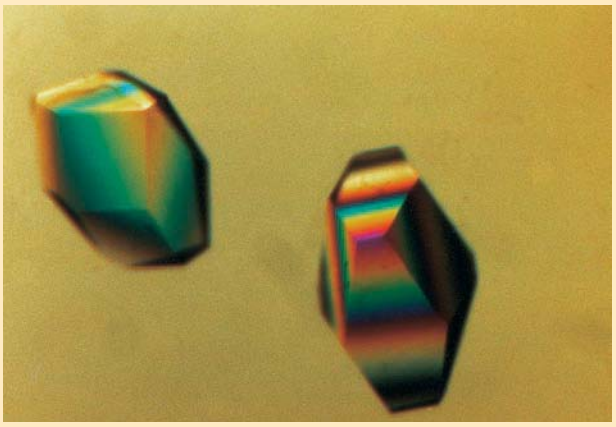


Figure B9.6 Crystals of lysozyme. The photograph was taken using polarizing filters to produce the color in the crystals. The actual size of these crystals is approximately 0.5 × 0.5 × 0.5 mm. (*Source:* Courtesy Fusao Takusagawa.)

protein. But visible light has wavelengths averaging about 500 nm (5000 Å). Thus, it is clearly impossible to resolve atoms with visible light. By contrast, x-rays have wavelengths of one to a few angstroms. For example, the characteristic x-rays emitted by excited copper atoms have a wavelength of 1.54 Å, which is ideal for high-resolution x-ray diffraction analysis of proteins.

In this chapter we will see protein structures at various levels of resolution. What is the reason for these differences in resolution? A protein crystal in which the protein molecules are relatively well ordered gives many diffraction spots far from the incident beam, that is, from the center of the detector. These spots are produced by x-rays with large diffraction angles (θ, see Figure B9.2). An electron-density map calculated from these diffraction spots from a relatively ordered crystal gives a high-resolution image of the diffracting molecule. On the other hand, a protein crystal whose molecules are relatively poorly arranged gives diffraction spots only near the center of the detector, resulting from x-rays with small diffraction angles. Such data produce a relatively low-resolution image of the molecule.

This relationship between resolution and diffraction angle is another consequence of Bragg's law $2d \sin \theta = \lambda$. Rearranging Bragg's equation, we find $d = \lambda/2 \sin \theta$. So we see that d, the distance between structural elements in the protein, is inversely related to sin θ. Therefore, the larger the distance between structural elements in the crystal, the smaller the angle of diffraction and the closer to the middle of the pattern the diffracted ray will fall. This is just another way of saying that low-resolution structure (with large distances between elements) gives rise to the pattern of spots near the middle of the diffraction pattern. By the same argument, high-resolution structure gives rise to spots near the periphery of the pattern because they diffract the x-rays at a large angle. When crystallographers can make crystals that are good enough to give this kind of high resolution, they can build a detailed model of the structure of the protein.

The proteins we are considering in this chapter are DNA-binding proteins. In many cases, investigators have prepared cocrystals of the protein and a double-stranded DNA fragment containing the target sequence recognized by the protein. These can reveal not only the shapes of the protein and DNA in the protein–DNA complex, but also the atoms that are involved in the protein–DNA interaction.

It is important to note that x-ray crystallography captures but one conformation of a molecule or collection of molecules. But proteins generally do not have just one possible conformation. They are dynamic molecules in constant motion and are presumably continuously sampling a range of different conformations. The particular conformation revealed by x-ray crystallography depends on the ligands that co-crystallize with the protein, and on the conditions used during crystallization.

Furthermore, a protein by itself may have a preferred conformation that seems incompatible with binding to a ligand, but its dynamic motions lead to other conformations that do permit ligand binding. For example, Max Perutz noted many years ago that the x-ray crystal structure of hemoglobin was not compatible with binding to its ligand, oxygen. Yet hemoglobin obviously does bind oxygen, and it does so by changing its shape enough to accommodate the ligand. Similarly, a DNA-binding protein by itself may prefer a conformation that cannot admit the DNA, but dynamic motions lead to another conformation that can bind the DNA, and the DNA traps the protein in that conformation.

the other hand, if the binding specificity had changed, the cells producing the recombinant repressor should have been immune to P22 infection. Actually, 434 and P22 do not infect *E. coli* cells, so the investigators used recombinant λ phages with the 434 and P22 immunity regions (λ_{imm}434 and λ_{imm}P22, respectively) in these tests. They found that the cells producing the altered 434 repressor *were* immune to infection by the λ phage with the P22 immunity region, but not to infection by the λ phage with the 434 immunity region.

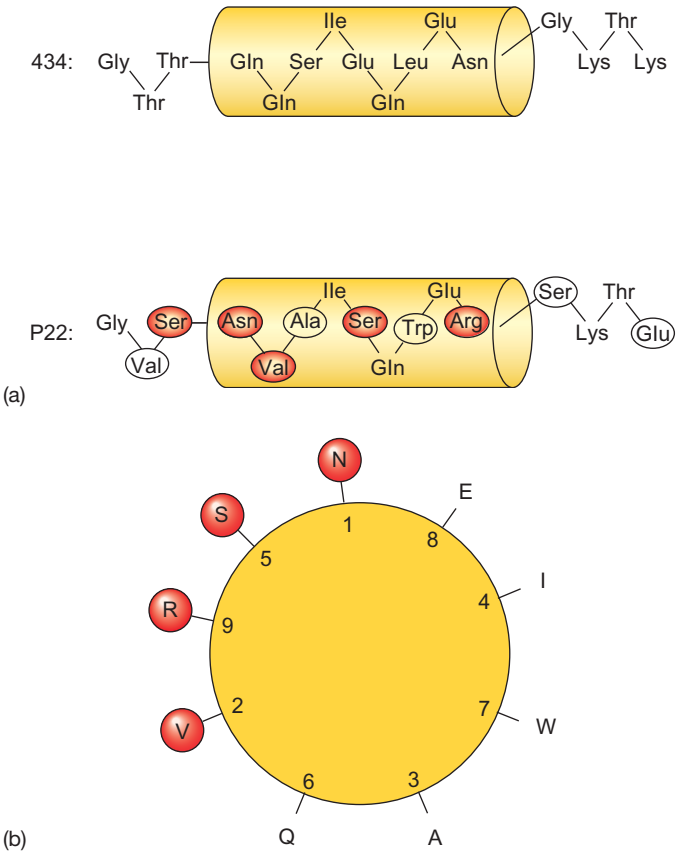


Figure 9.3 The recognition helices of two λ-like phage repressors. **(a)** Key amino acids in the recognition helices of two repressors. The amino acid sequences of the recognition helices of the 434 and P22 repressors are shown, along with a few amino acids on either side. Amino acids that differ between these two proteins are circled in the P22 diagram; these are more likely to contribute to differences in specificity. Furthermore, the amino acids on the side of the helix that faces the DNA are most likely to be involved in DNA binding. These, along with one amino acid in the turn just before the helix (red), were changed to alter the binding specificity of the protein. **(b)** The recognition helix of the P22 repressor viewed on end. The numbers represent the positions of the amino acids in the protein chain. The left-hand side of the helix faces toward the DNA, so the amino acids on that side are more likely to be important in binding. Those that differ from amino acids in corresponding positions in the 434 repressor are circled in red. (*Source: (b)* Adapted from Wharton, R.P. and M. Ptashne, Changing the binding specificity of a repressor by redesigning an alpha-helix. *Nature* 316:602, 1985.)

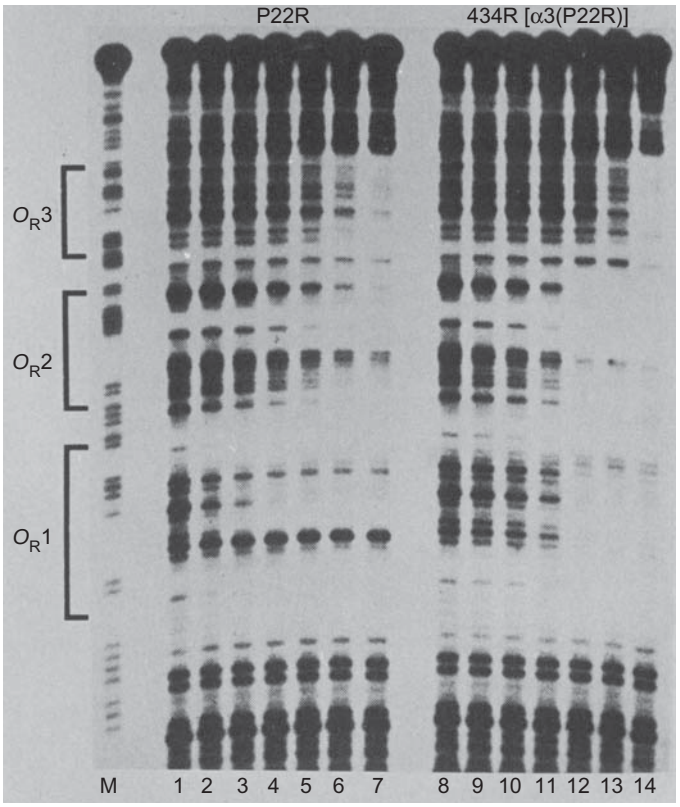


Figure 9.4 DNase footprinting with the recombinant 434 repressor. Wharton and Ptashne performed DNase footprinting with end-labeled P22 phage O_R and either P22 repressor (P22R, lanes 1–7) or the 434 repressor with five amino acids in the recognition helix (α-helix 3) changed to match those in the phage P22 recognition helix (434R[α3(P22R)], lanes 8–14). The two sets of lanes contained increasing concentrations of the respective repressors (0 M in lanes 1 and 8, and ranging from 7.6×10^{-10} M to 1.1×10^{-8} M in lanes 2–7 and from 5.2×10^{-9} M to 5.6×10^{-7} M in lanes 8–14). The marker lane (M) contained the A + G reaction from a sequencing procedure. The positions of all three rightward operators are indicated with brackets at left. (*Source:* Wharton, R.P. and M. Ptashne, Changing the binding specificity of a repressor by redesigning an alpha-helix. *Nature* 316 (15 Aug 1985), f. 3, p. 603. © Macmillan Magazines Ltd.)

To check these results, Wharton and Ptashne measured DNA binding in vitro by DNase footprinting (Chapter 5). They found that the purified recombinant repressor could make a "footprint" in the P22 operator, just as the P22 repressor can (Figure 9.4). In control experiments (not shown) they demonstrated that the recombinant repressor could no longer make a footprint in the 434 operator. Thus, the binding specificity really had been altered by these five amino acid changes. In further experiments, Ptashne and colleagues showed that the first four of these amino acids were necessary and sufficient for either binding activity. That is, if the repressor had TQQE (threonine, glutamine, glutamine, glutamate) in its recognition helix, it would bind to the 434 operator. On the other hand, if it had SNVS (serine, asparagine, valine, serine), it would bind to the P22 operator.

What if Wharton and Ptashne had not tried to *change* the specificity of the repressor, but just to *eliminate* it? They could have identified the amino acids in the repressor that were probably important to specificity, then changed them to other amino acids chosen at random and shown that this recombinant 434 repressor could no longer bind to its operator. If that is all they had done, they could have said that the results were consistent with the hypothesis that the altered amino acids are directly involved in binding. But an alternative explanation would remain: These amino acids could simply be important to the overall three-dimensional shape of the repressor protein, and changing them changed this shape and therefore indirectly prevented binding. By contrast, changing specificity by changing amino acids is

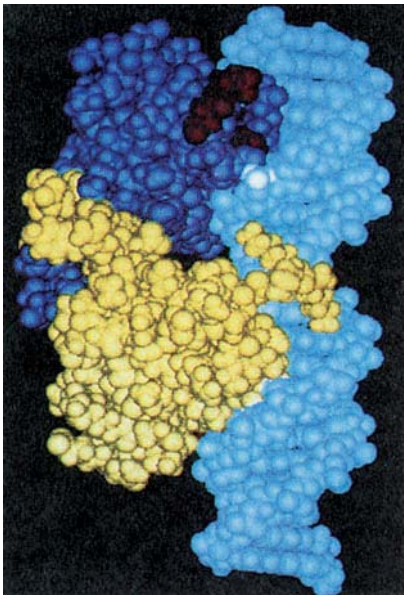


Figure 9.5 Computer model of the λ repressor dimer binding to λ operator (O_R2). The DNA double helix (light blue) is at right. The two monomers of the repressor are in dark blue and yellow. The helix-turn-helix motif of the upper monomer (dark red and blue) is inserted into the major groove of the DNA. The arm of the lower monomer reaches around to embrace the DNA. (*Source:* Hochschild, A., N. Irwin, and M. Ptashne, Repressor structure and the mechanism of positive control. *Cell* 32 (1983) p. 322. Reprinted by permission of Elsevier Science. Photo by Richard Feldman.)

strong evidence for the direct involvement of these amino acids in binding.

In a related x-ray crystallographic study, Ptashne and coworkers showed that the λ repressor has an amino-terminal arm not found in the repressors of the 434 and P22 phages. This arm contributes to the repressor's binding to the λ operator by embracing the operator. Figure 9.5 shows a computer model of a dimer of λ repressor interacting with λ operator. In the repressor monomer at the top, the helix-turn-helix motif is visible projecting into the major groove of the DNA. At the bottom, we can see the arm of the other repressor monomer reaching around to embrace the DNA.

Cro also uses a helix-turn-helix DNA binding motif and binds to the same operators as the λ repressor, but it has the exact opposite affinity for the three different operators in a set (Chapter 8). That is, it binds first to O_R3 and last to O_R1, rather than vice versa. Therefore, by changing amino acids in the recognition helices, one ought to be able to identify the amino acids that give Cro and the λ repressor their different binding specificities. Ptashne and his coworkers accomplished this task and found that amino acids 5 and 6 in the recognition helices are especially important, as is the amino-terminal arm in the λ repressor. When these workers altered base pairs in the operators, they discovered that the base pairs critical to discriminating between O_R1 and O_R3 are at position 3, to which Cro is more sensitive, and at positions 5 and 8, which are selective for repressor binding.

SUMMARY The repressors of the λ-like phages have recognition helices that fit sideways into the major groove of the operator DNA. Certain amino acids on the DNA side of the recognition helix make specific contact with bases in the operator, and these contacts determine the specificity of the protein–DNA interactions. In fact, changing these amino acids can change the specificity of the repressor. The λ repressor itself has an extra motif not found in the other repressors, an amino-terminal arm that aids binding by embracing the DNA. The λ repressor and Cro share affinity for the same operators, but they have microspecificities for O_R1 or O_R3, determined by interactions between different amino acids in the recognition helices of the two proteins and different base pairs in the two operators.

High-Resolution Analysis of λ Repressor–Operator Interactions

Steven Jordan and Carl Pabo wished to visualize the λ repressor–operator interaction at higher resolution than previous studies allowed. They were able to achieve a resolution of 2.5 Å by making excellent cocrystals of a repressor fragment and an operator fragment. The repressor fragment encompassed residues 1–92, which included all of the DNA-binding domain of the protein. The operator fragment (Figure 9.6) was 20 bp long and contained one complete site to which the repressor dimer attached. That is, it had two half-sites, each of which bound to a repressor monomer. Such use of partial molecules is a common trick employed by x-ray crystallographers to make better crystals than they can obtain with whole proteins or whole DNAs. In this case, because the primary goal was to elucidate the structure of the interface between the repressor and the operator, the protein and DNA fragments were probably just as useful as the whole protein and DNA because they contained the elements of interest.

General Structural Features Figure 9.2, used at the beginning of this chapter to illustrate the fit between λ repressor and operator, is based on the high-resolution model from the

Figure 9.6 The operator fragment used to prepare operator–repressor cocrystals. This 20-mer contains the two λ O_L1 half-sites, each of which binds a monomer of repressor. The half-sites are included within the 17-bp region in boldface; each half-site contains 8 bp, separated by a G–C pair in the middle (9). The half-site on the left has a consensus sequence; that on the right deviates somewhat from the consensus. The base pairs of the consensus half-site are numbered 1–8; those in the other half-site are numbered 1′–8′.

Figure 9.7 Geometry of the λ repressor–operator complex. The DNA (blue) is bound to the repressor dimer, whose monomers are depicted in yellow and purple. The recognition helix of each monomer is shown in red and labeled 3 and 3′. (*Source:* Adapted from Jordan, S.R. and C.O. Pabo, Structure of the lambda complex at 2.5 Å resolution. Details of the repressor–operator interaction. *Science* 242:895, 1988.)

Jordan and Pabo analysis we are now considering. Figure 9.7, a more detailed representation of the same model, reveals several general aspects of the protein–DNA interaction. First of all, of course, we can see the recognition helices (3 and 3′, red) of each repressor monomer nestled into the DNA major grooves in the two half-sites. We can also see how helices 5 and 5′ approach each other to hold the two monomers together in the repressor dimer. Finally, note that the DNA is similar in shape to the standard B-form of DNA. We can see a bit of bending of the DNA, especially at the two ends of the DNA fragment, as it curves around the repressor dimer, but the rest of the helix is relatively straight.

Interactions with Bases Figure 9.8 shows the details of the interactions between amino acids in a repressor monomer and bases in one operator half-site. The crucial amino acids participating in these interactions are glutamine 33 (Gln 33), glutamine 44 (Gln 44), serine 45 (Ser 45), lysine 4 (Lys 4), and asparagine 55 (Asn 55). Figure 9.8a is a stereo view of the interactions, where α-helices 2 and 3 are represented by bold lines. The recognition helix (3) is almost perpendicular to the plane of the paper, so the helical polypeptide backbone looks like a bumpy circle. The key amino acid side chains are shown making hydrogen bonds (dashed lines) to the DNA and to one another.

Figure 9.8b is a schematic diagram of the same amino acid/DNA interactions. It is perhaps easier to see the hydrogen bonds in this diagram. We see that three of the important bonds to DNA bases come from amino acids in the recognition helix. In particular, Gln 44 makes two hydrogen bonds to adenine-2, and Ser 45 makes one hydrogen bond to guanine-4. Figure 9.8c depicts these hydrogen bonds in detail and also clarifies a point made in parts (a) and (b) of the figure: Gln 44 also makes a hydrogen bond to Gln 33, which in turn is hydrogen-bonded to the phosphate preceding base pair number 2. This is an example of a **hydrogen bond network,** which involves three or more entities (e.g., amino acids, bases, or DNA backbone). The participation of Gln 33 is critical. By bridging between the DNA backbone and Gln 44, it positions Gln 44 and the rest of the recognition helix to interact optimally with the operator. Thus, even though Gln 33 resides at the beginning of helix 2, rather than on the recognition helix, it plays an important role in protein–DNA binding. To underscore the importance of this glutamine, we note that it also appears in the same position in the 434 phage repressor and plays the same role in interactions with the 434 operator, which we will examine later in this chapter.

Serine 45 also makes an important hydrogen bond with a base pair, the guanine of base pair number 4. In addition, the methylene (CH_2) group of this serine approaches the methyl group of the thymine of base pair number 5 and participates in a hydrophobic interaction that probably also includes the methyl group of Ala 49. Such hydrophobic interactions involve nonpolar groups like methyl and methylene, which tend to come together to escape the polar environment of the water solvent, much as oil droplets coalesce to minimize their contact with water. Indeed, hydrophobic literally means "water-fearing."

The other hydrogen bonds with base pairs involve two other amino acids that are not part of the recognition helix. In fact, these amino acids are not part of any helix: Asn 55 lies in the linker between helices 3 and 4, and Lys 4 is on the arm that reaches around the DNA. Here again we see an example of a hydrogen bond network, not only between amino acid and base, but between two amino acids. Figure 9.8c makes it particularly clear that these two amino acids each form hydrogen bonds to the guanine of base pair number 6, and also to each other. Such networks add considerably to the stability of the whole complex.

Amino Acid/DNA Backbone Interactions We have already seen one example of an amino acid (Gln 33) that forms a hydrogen bond with the DNA backbone (the phosphate between base pairs 1 and 2). However, this is only one of five such interactions in each half-site. Figure 9.9 portrays these interactions in the consensus half-site, which involve five different amino acids, only one of which (Asn 52) is in the recognition helix. The dashed lines represent hydrogen bonds from the NH groups of the peptide backbone, rather than from the amino acid side chains.

One of these hydrogen bonds, involving the peptide NH at Gln 33, is particularly interesting because of an electrostatic contribution of helix 2 as a whole. To appreciate this, recall from Chapter 3 that all the C═O bonds in a protein α-helix point in one direction. Because each of these bonds is polar, with a partial negative charge on the oxygen and a partial positive charge on the carbon, the whole α-helix has

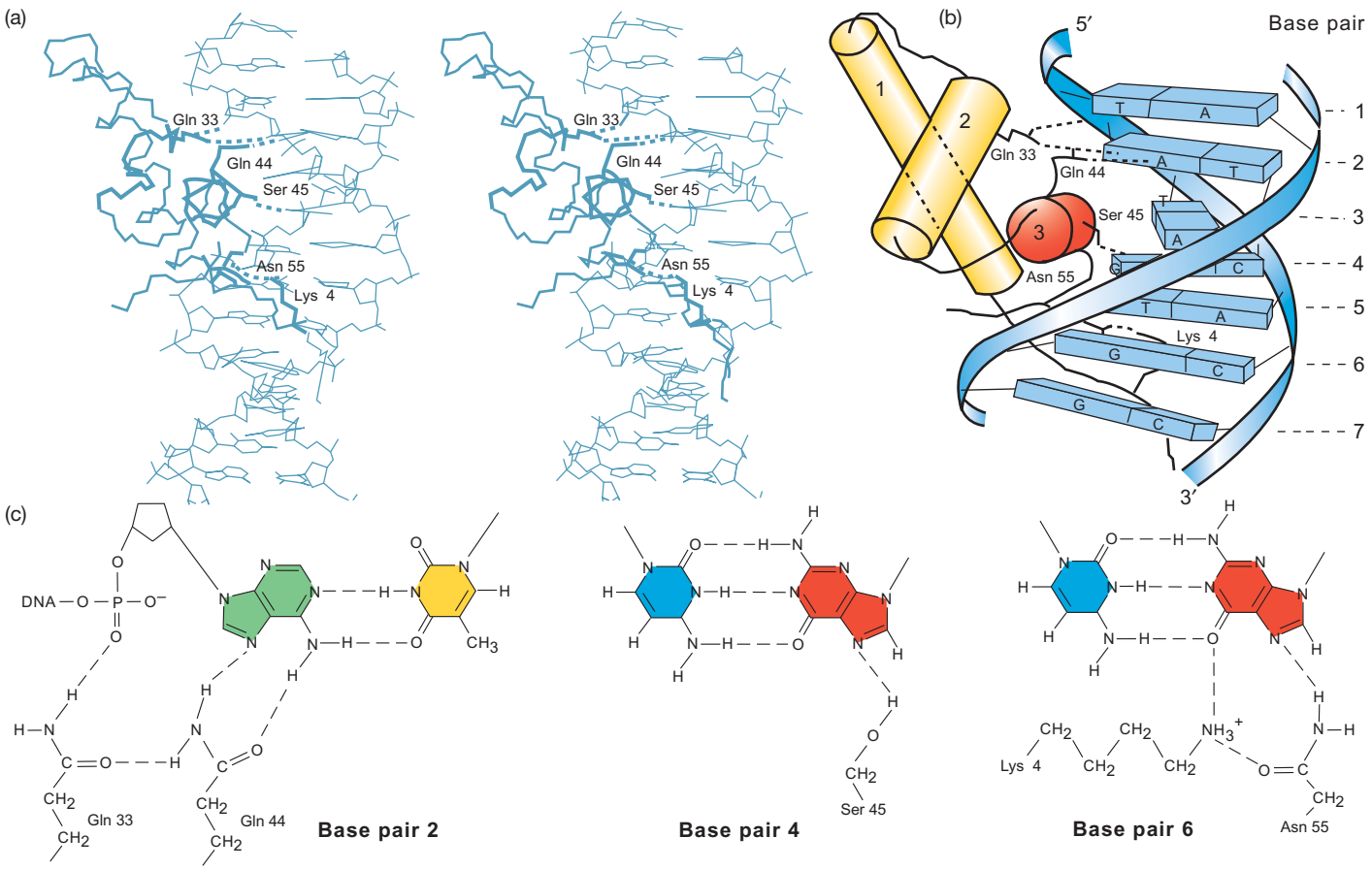


Figure 9.8 Hydrogen bonds between λ repressor and base pairs in the major groove of the operator. (a) Stereo diagram of the complex, with the DNA double helix on the right and the amino terminal part of the repressor monomer on the left. α-Helices 2 and 3 are rendered in bold lines, with the recognition helix almost perpendicular to the plane of the paper. Hydrogen bonds are represented by dashed lines. **(b)** Schematic diagram of the hydrogen bonds shown in panel (a). Only the important amino acid side chains are shown. The base pairs are numbered at right. **(c)** Details of the hydrogen bonds. Structures of the key amino acid side chains and bases are shown, along with the hydrogen bonds in which they participate. (*Source:* From Jordan, S.R. and C.O. Pabo, Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science* 242:896, 1988. Copyright © 1988 AAAS. Reprinted with permission from AAAS.)

a considerable polarity, with practically a full net positive charge at the amino terminus of the helix. This end of the helix will therefore have a natural affinity for the negatively charged DNA backbone. Now look again at Figure 9.9 and notice that the amino end of helix 2, where Gln 33 is located, points directly at the DNA backbone. This maximizes the electrostatic attraction between the positively charged amino end of the α-helix and the negatively charged DNA and stabilizes the hydrogen bond between the peptide NH of Gln 33 and the phosphate group in the DNA backbone.

Other interactions involve hydrogen bonds between amino acid side chains and DNA backbone phosphates. For example, Lys 19 and Asn 52 both form hydrogen bonds with phosphate P_B. The amino group of Lys 26 carries a full positive charge. Although it may be too far away from the DNA backbone to interact directly with a phosphate, it may contribute to the general affinity between protein and DNA. The large number of amino acid/DNA phosphate contacts suggests that these interactions play a major role in the stabilization of the protein–DNA complex. Figure 9.9 also shows the position of the side chain of Met 42. It probably forms a hydrophobic interaction with three carbon atoms on the deoxyribose between P_C and P_D.

Confirmation of Biochemical and Genetic Data Before the detailed structure of the repressor–operator complex was known, we already had predictions from biochemical and genetic experiments about the importance of certain repressor amino acids and operator bases. In almost all cases, the structure confirms these predictions.

First, ethylation of certain operator phosphates interfered with repressor binding. Hydroxyl radical footprinting had also implicated these phosphates in repressor binding. Now we see that these same phosphates (five per half-site) make important contacts with repressor amino acids in the cocrystal.

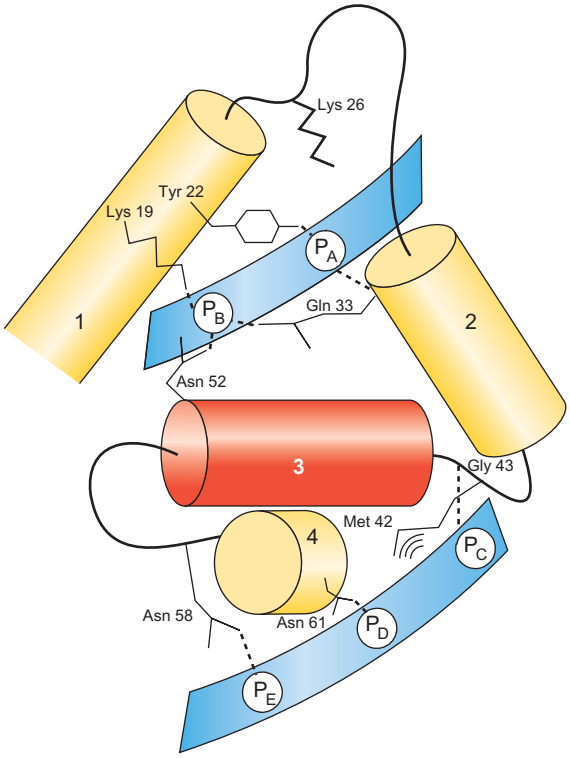


Figure 9.9 Amino acid/DNA backbone interactions. α-Helices 1–4 of the λ repressor are shown, along with the phosphates (P_A–P_E) that are involved in hydrogen bonds with the protein. This diagram is perpendicular to that in Figure 9.8. The side chains of the important amino acids are shown. The two dashed lines denote hydrogen bonds between peptide NH groups and phosphates. Concentric arcs denote a hydrophobic interaction. (*Source:* Adapted from Jordan S.R. and C.O. Pabo, Structure of the lambda complex at 2.5 Å resolution: Details of the repressor–operator interactions. *Science* 242:897, 1988.)

Second, methylation protection experiments had predicted that certain guanines in the major groove would be in close contact with repressor. The crystal structure now shows that all of these are indeed involved in repressor binding. One major-groove guanine actually became more sensitive to methylation on repressor binding, and this guanine (G8′, Figure 9.6) is now seen to have an unusual conformation in the cocrystal. Base pair 8′ is twisted more than any other on its horizontal axis, and the spacing between this base pair and the next is the widest. This unusual conformation could open guanine 8′ up to attack by the methylating agent DMS. Also, adenines were not protected from methylation in previous experiments. This makes sense because adenines are methylated on N3, which resides in the minor groove. Because no contacts between repressor and operator occur in the minor groove, repressor cannot protect adenines from methylation.

Third, DNA sequence data had shown that the A–T base pair at position 2 and the G–C base pair at position 4 (Figure 9.8) were conserved in all 12 half-sites of the operators O_R and O_L. The crystal structure shows why these base pairs are so well conserved: They are involved in important contacts with the repressor.

Fourth, genetic data had shown that mutations in certain amino acids destabilized repressor–operator interaction, whereas other changes in repressor amino acids actually enhanced binding to the operator. Almost all of these mutations can be explained by the cocrystal structure. For example, mutations in Lys 4 and Tyr 22 were particularly damaging, and we now see (Figures 9.8 and 9.9) that both these amino acids make strong contacts with the operator: Lys 4 with guanine-6 (and with Asn 55) and Tyr 22 with P_A. As an example of a mutation with a positive effect, consider the substitution of lysine for Glu 34. This amino acid is not implicated by the crystal structure in any important bonds to the operator, but a lysine in this position could rotate so as to form a salt bridge with the phosphate before P_A (Figure 9.9) and thus enhance protein–DNA binding. This salt bridge would involve the positively charged ε-amino group of the lysine and the negatively charged phosphate.

SUMMARY The cocrystal structure of a λ repressor fragment with an operator fragment shows many details about how the protein and DNA interact. The most important contacts occur in the major groove, where amino acids make hydrogen bonds with DNA bases and with the DNA backbone. Some of these hydrogen bonds are stabilized by hydrogen-bond networks involving two amino acids and two or more sites on the DNA. The structure derived from the cocrystal is in almost complete agreement with previous biochemical and genetic data.

High-Resolution Analysis of Phage 434 Repressor–Operator Interactions

Harrison, Ptashne, and coworkers used x-ray crystallography to perform a detailed analysis of the interaction between phage 434 repressor and operator. As in the λ cocrystal structure, the crystals they used for this analysis were not composed of full-length repressor and operator, but fragments of each that contained the interaction sites. As a substitute for the repressor, they used a peptide containing the first 69 amino acids of the protein, including the helix-turn-helix DNA-binding motif. For the operator, they used a synthetic 14-bp DNA fragment that contains the repressor-binding site. These two fragments presumably bound together as the intact molecules would, and the complex could be crystallized relatively easily. We will focus here on concepts that were not clearly demonstrated by the λ repressor–operator studies.

Contacts with Base Pairs Figure 9.10 summarizes the contacts between the side chains of Gln 28, Gln 29, and Gln 33, all in the recognition helix (α3) of the 434 repressor. Starting at the bottom of the figure, note the two possible

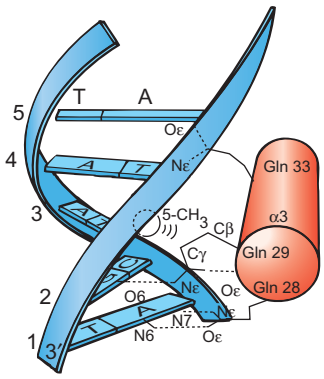


Figure 9.10 Detailed model of interaction between recognition helix amino acid side chains and one 434 operator half-site. Hydrogen bonds are represented by dashed lines. The van der Waals interaction between the Gln 29 side chain and the 5-methyl group of the thymine paired to adenine 3 is represented by concentric arcs. (*Source:* Adapted from Anderson, J. E., M. Ptashne, and S. C. Harrison, Structure of the repressor–operator complex of bacteriophage 434. *Nature* 326:850, 1987.)

hydrogen bonds (represented by dashed lines) between the Oε and Nε of Gln 28 and the N6 and N7 of adenine 1. Next, we see that a possible hydrogen bond between the Oε of Gln 29 and the protein backbone NH of the same amino acid points the Nε of this amino acid directly at the O6 of the guanine in base pair 2 of the operator, which would allow a hydrogen bond between this amino acid and base. Note also the potential van der Waals interactions (represented by concentric arcs) between Cβ and Cγ of Gln 29 and the 5-methyl group of the thymine in base pair 3. Such van der Waals interactions can be explained roughly as follows: Even though all the groups involved are nonpolar, at any given instant they have a very small dipole moment due to random fluctuations in their electron clouds. These small dipole moments can cause a corresponding opposite polarity in a very close neighbor. The result is an attraction between the neighboring groups.

SUMMARY X-ray crystallography of a phage 434 repressor-fragment/operator-fragment complex shows probable hydrogen bonding between three glutamine residues in the recognition helix and three base pairs in the repressor. It also reveals a potential van der Waals contact between one of these glutamines and a base in the operator.

Effects of DNA Conformation The contacts between the repressor and the DNA backbone require that the DNA double helix curve slightly. Indeed, higher-resolution crystallography studies by Harrison, Ptashne, and colleagues show that the DNA does curve this way in the DNA–protein complex (Figure 9.11); we do not know yet whether the DNA bend preexists in this DNA region or whether it is induced by repressor binding. In either case, the base sequence of the operator plays a role by facilitating this bending. That is, some DNA sequences are easier to bend in a given way than others, and the 434 operator sequence is optimal for the bend it must make to fit the repressor. We will discuss this general phenomenon in more detail later in this chapter.

Another notable feature of the conformation of the operator DNA is the compression of the DNA double helix between base pairs 7 and 8, which lie between the two half-sites of the operator. This compression amounts to an overwinding of 3 degrees between base pairs 7 and 8, or 39 degrees, compared with the normal 36 degrees helical twist between base pairs. Notice the narrowness of the minor groove at center right in Figure 9.11b, compared to Figure 9.11a. The major grooves on either side are wider than normal, due to a compensating underwinding of that DNA. Again, the base sequence at this point is optimal for assuming this conformation.

SUMMARY The x-ray crystallography analysis of the partial phage 434 repressor–operator complex shows that the DNA deviates significantly from its normal regular shape. It bends somewhat to accommodate the necessary base/amino acid contacts. Moreover, the central part of the helix, between the two half-sites, is wound extra tightly, and the outer parts are wound more loosely than normal. The base sequence of the operator facilitates these departures from normal DNA shape.

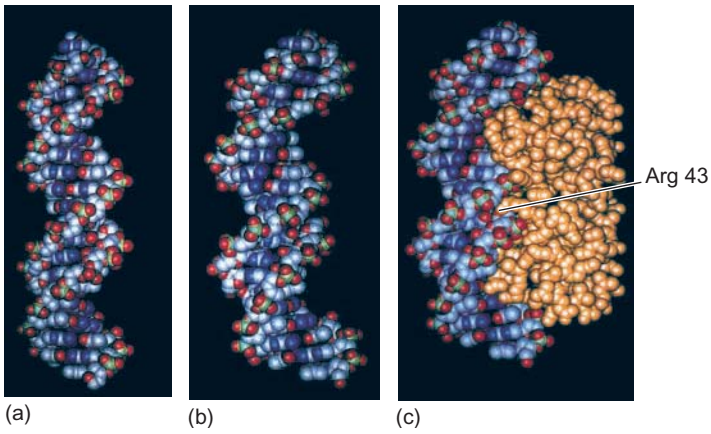


Figure 9.11 Space-filling computer model of distorted DNA in the 434 repressor–operator complex. (a) Standard B-DNA. **(b)** Shape of the operator-containing 20-mer in the repressor–operator complex with the protein removed. Note the overall curvature, and the narrowness of the minor groove at center right. **(c)** The repressor–operator complex, with the repressor in orange. Notice how the DNA conforms to the shape of the protein to promote intimate contact between the two. The side chain of Arg 43 can be seen projecting into the minor groove of the DNA near the center of the model. (*Source:* Aggarwal et al., Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. *Science* 242 (11 Nov 1988) f. 3b, f. 3c, p. 902. © AAAS.)

Genetic Tests of the Model If the apparent contacts we have seen between repressor and operator are important, mutations that change these amino acids or bases should reduce or abolish DNA–protein binding. Alternatively, we might be able to mutate the operator so it does not fit the repressor, then make a compensating mutation in the repressor that restores binding. Also, if the unusual shape assumed by the operator is important, mutations that prevent it from taking that shape should reduce or abolish repressor binding. As we will see, all those conditions have been fulfilled.

To demonstrate the importance of the interaction between Gln 28 and A1, Ptashne and colleagues changed A1 to a T. This destroyed binding between repressor and operator, as we would expect. However, this mutation could be suppressed by a mutation at position 28 of the repressor from Gln to Ala. Figure 9.10 reveals the probable explanation: The two hydrogen bonds between Gln 28 and A1 can be replaced by a van der Waals contact between the methyl groups on Ala 28 and T1. The importance of this contact is underscored by the replacement of T1 with a uracil, which does not have a methyl group, or 5-methylcytosine (5MeC), which does. The U-substituted operator does not bind the repressor with Ala 28, but the 5MeC-substituted operator does. Thus, the methyl group is vital to interactions between the mutant operator and mutant repressor, as predicted on the basis of the van der Waals contact.

We strongly suspect that the overwinding of the DNA between base pairs 7 and 8 is important in repressor–operator interaction. If so, substituting G–C or C–G base pairs for the A–T and T–A pairs at positions 6–9 should decrease repressor–operator binding, because G–C pairs do not readily allow the overwinding that is possible with A–T pairs. As expected, repressor did not bind well to operators with G–C or C–G base pairs in this region. This failure to bind well did not prove that overwinding exists, but it was consistent with the overwinding hypothesis.

SUMMARY The contacts between the phage 434 repressor and operator predicted by x-ray crystallography can be confirmed by genetic analysis. When amino acids or bases predicted to be involved in interaction are altered, repressor–operator binding is inhibited. Furthermore, binding is also inhibited when the DNA is mutated so it cannot as readily assume the shape it has in the repressor–operator complex.

9.2 The *trp* Repressor

The *trp* repressor is another protein that uses a helix-turn-helix DNA-binding motif. However, recall from Chapter 7 that the aporepressor (the protein without the tryptophan corepressor) is not active. Paul Sigler and colleagues used x-ray crystallography of *trp* repressor and aporepressor to point out the subtle but important difference that tryptophan makes. The crystallography also sheds light on the way the *trp* repressor interacts with its operator.

The Role of Tryptophan

Here is a graphic indication that tryptophan affects the shape of the repressor: When you add tryptophan to crystals of aporepressor, the crystals shatter! When the tryptophan wedges itself into the aporepressor to form the repressor, it changes the shape of the protein enough to break the lattice forces holding the crystal together.

This raises an obvious question: What moves when free tryptophan binds to the aporepressor? To understand the answer, it helps to visualize the repressor as illustrated in Figure 9.12. The protein is actually a dimer of identical subunits, but these subunits fit together to form a three-domain structure. The central domain, or "platform," comprises the A, B, C, and F helices of each monomer, which are grouped together on the right, away from the DNA. The other two domains, found on the left close to the DNA, are the D and E helices of each monomer.

Now back to our question: What moves when we add tryptophan? The platform apparently remains stationary, whereas the other two domains tilt, as shown in Figure 9.12. The recognition helix in each monomer is helix E, and we can see an obvious shift in its position when tryptophan binds. In the top monomer, it shifts from a somewhat downward orientation to a position in which it points directly into the major groove of the operator. In this position, it is ideally situated to make contact with (or "read") the DNA, as we will see.

Sigler refers to these DNA-reading motifs as *reading heads,* likening them to the heads in the hard drive of a computer. In a computer, the reading heads can assume two positions: engaged and reading the drive, or disengaged and away from the drive. The *trp* repressor works the same way. When tryptophan is present, it inserts itself between the platform and each reading head, as illustrated in Figure 9.12, and forces the reading heads into the best position (transparent helices D and E) for fitting into the major groove of the operator. On the other hand, when tryptophan dissociates from the aporepressor, the gap it leaves allows the reading heads to fall back toward the central platform and out of position to fit with the operator (gray helices D and E).

Figure 9.13a shows a closer view of the environment of the tryptophan in the repressor. It is a hydrophobic pocket that is occupied by the side chain of a hydrophobic amino acid (sometimes tryptophan) in almost all comparable helix-turn-helix proteins, including the λ repressor, Cro, and CAP. However, in these other proteins the hydrophobic amino acid is actually part of the protein chain, not a free amino acid, as in the *trp* repressor. Sigler likened the arrangement of the tryptophan between Arg 84 and Arg 54

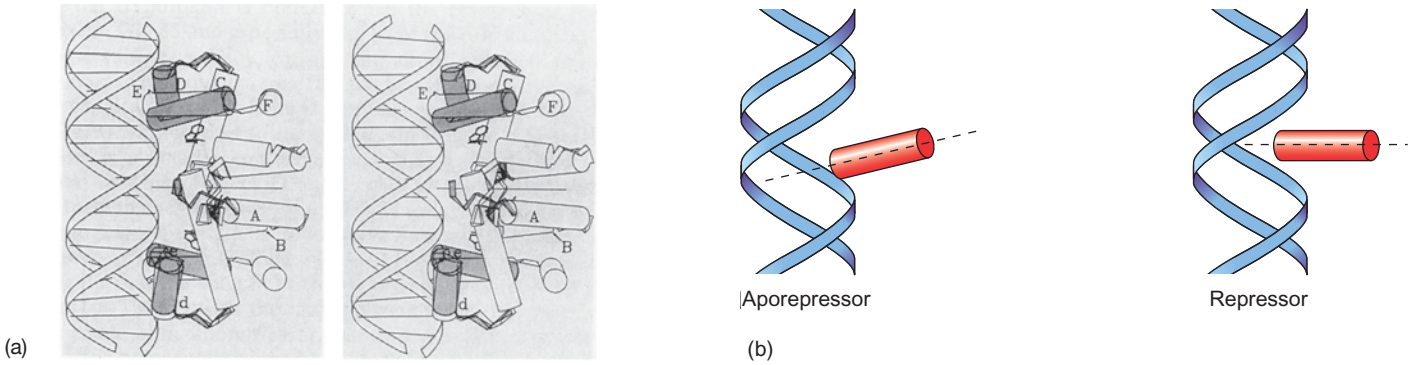


Figure 9.12 Comparison of the fit of *trp* repressor and aporepressor with *trp* operator. (a) Stereo diagram. The helix-turn-helix motifs of both monomers are shown in the positions they assume in the repressor (transparent) and aporepressor (dark). The position of tryptophan in the repressor is shown (black polygons). Note that the recognition helix (helix E) in the aporepressor falls back out of ideal position for inserting into the major groove of the operator DNA. The two almost identical drawings constitute a stereo presentation that allows you to view this picture in three dimensions. To get this 3-D effect, use a stereo viewer, or alternatively, hold the picture 1–2 ft in front of you and let your eyes relax as they would when you are staring into the distance or viewing a "magic eye" picture. After a few seconds, the two images should fuse into one in the center, which appears in three dimensions. This stereo view gives a better appreciation for the fit of the recognition helix and the major groove of the DNA, but if you cannot get the 3-D effect, just look at one of the two pictures. **(b)** Simplified (nonstereo) diagram comparing the positions of the recognition helix (red) of the aporepressor (left) and the repressor (right) with respect to the DNA major groove. Notice that the recognition helix of the repressor points directly into the major groove, whereas that of the aporepressor points more downward. The dashed line emphasizes the angle of the recognition helix in each drawing.

to a salami sandwich, in which the flat tryptophan is the salami. When it is removed, as in Figure 9.13b, the two arginines come together as the pieces of bread would when you remove the salami from a sandwich. This model has implications for the rest of the molecule, because Arg 54 is on the surface of the central platform of the repressor dimer, and Arg 84 is on the facing surface of the reading head. Thus, inserting the tryptophan between these two arginines pushes the reading head away from the platform and points it toward the major groove of the operator, as we saw in Figure 9.12.

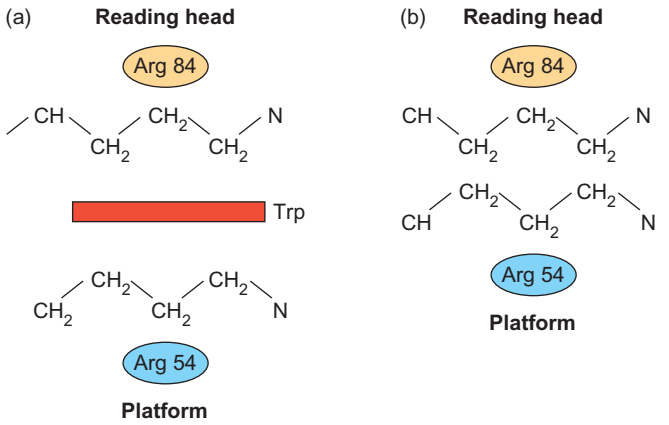


Figure 9.13 Tryptophan-binding site in the *trp* repressor. (a) Environment surrounding the tryptophan (Trp) in the *trp* repressor. Notice the positions of Arg 84 above and Arg 54 below the tryptophan side chain (red). **(b)** The same region in the aporepressor, without tryptophan. Notice that the Arg side chains have moved together to fill the gap left by the absent tryptophan.

SUMMARY The *trp* repressor requires tryptophan to force the recognition helices of the repressor dimer into the proper position for interacting with the *trp* operator.

9.3 General Considerations on Protein–DNA Interactions

What contributes to the specificity of binding between a protein and a specific stretch of DNA? The examples we have seen so far suggest two answers: (1) specific interactions between bases and amino acids; and (2) the ability of the DNA to assume a certain shape, which also depends on the DNA's base sequence (a phenomenon Sigler calls "indirect readout"). These two possibilities are clearly not mutually exclusive, and both apply to many of the same protein–DNA interactions.

Hydrogen Bonding Capabilities of the Four Different Base Pairs

We have seen that different DNA-binding proteins depend to varying extents on contacts with the bases in the DNA. To the extent that they "read" the sequence of bases, one can

ask, What exactly do they read? After all, the base pairs do not open up, so the DNA-binding proteins have to sense the differences among the bases in their base-paired condition. And they have to make base-specific contacts with these base pairs, either through hydrogen bonds or van der Waals interactions. Let us examine further the hydrogen-bonding potentials of the four different base pairs.

Consider the DNA double helix in Figure 9.14a. If we were to rotate the DNA 90 degrees so that it is sticking out of the page directly at us, we would be looking straight down the helical axis. Now consider one base pair of the DNA in this orientation, as pictured in Figure 9.14b. The major groove is on top, and the minor groove is below. A DNA-binding protein can approach either of these grooves to interact with the base pair. As it does so, it "sees" four possible contours in each groove, depending on whether the base pair is a T–A, A–T, C–G, or G–C pair.

Figure 9.14c presents two of these contours from both the major and minor groove perspectives. At the very bottom we see line diagrams (Figure 9.14d) that summarize what the protein encounters in both grooves for an A–T and a G–C base pair. Hydrogen bond acceptors (oxygen and nitrogen atoms) are denoted "Acc," and hydrogen bond donors (hydrogen atoms) are denoted "Don." The major and minor grooves lie above and below the horizontal lines, respectively. The lengths of the vertical lines represent the relative distances that the donor or acceptor atoms project away from the helical axis toward the outside of the DNA groove. We can see that the A–T and G–C base pairs present very different profiles to the outside world, especially in the major groove. The difference between a pyrimidine–purine pair and the purine–pyrimidine pairs shown here would be even more pronounced.

These hydrogen-bonding profiles assume direct interactions between base pairs and amino acids. However, other possibilities exist. There is indirect readout, in which amino acids "read" the shape of the DNA backbone, either by direct hydrogen bonding or by forming salt bridges. Amino acids and bases can also interact indirectly through hydrogen bonds to an intervening water molecule, but these "indirect interactions" are no less specific than direct ones.

SUMMARY The four different base pairs present four different hydrogen-bonding profiles to amino acids approaching either the major or minor DNA groove.

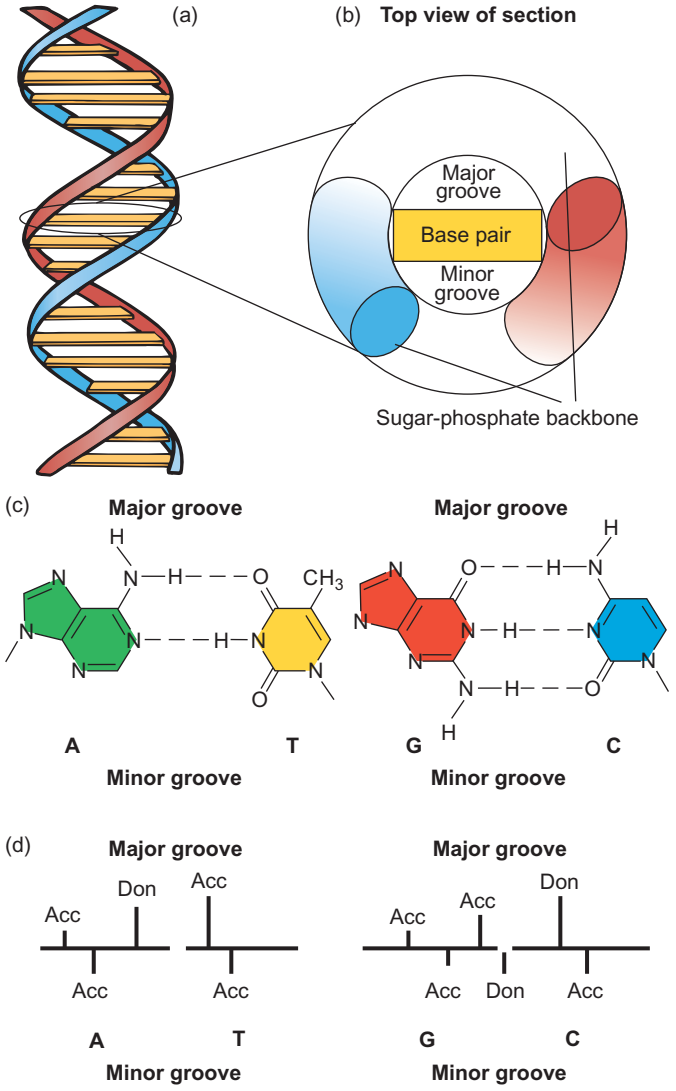


Figure 9.14 Appearance of base pairs in the major and minor grooves of DNA. (a) Standard B-form DNA, with the two backbones in red and blue, and the base pairs in yellow. **(b)** Same DNA molecule seen from the top. Notice the wider opening to the major groove (top), compared with the minor groove (bottom). **(c)** Structural formulas of the two base pairs. Again, the major groove is on top, and the minor groove on the bottom. **(d)** Line diagrams showing the positions of hydrogen bond acceptors (Acc) and donors (Don) in the major and minor grooves. For example, reading left to right, the major groove of the T–A pair has an acceptor (the N–7 in the ring of the adenine), then a donor (the NH_2 of the adenine), then an acceptor (the C=O of the thymine). The relative horizontal positions of these groups are indicated by the point of intersection with the vertical lines. The relative vertical positions are indicated by the lengths of the vertical lines. The two base pairs present different patterns of donors and acceptors in both major and minor grooves, so they are perceived differently by proteins approaching from the outside. By inverting these diagrams left-to-right, you can see that T–A and C–G pairs would present still different patterns.
(*Source:* Adapted from R. Schleif, DNA binding by proteins. *Science* 241:1182–3, 1988.)

The Importance of Multimeric DNA-Binding Proteins

Robert Schleif noted that the target sites for DNA-binding proteins are usually symmetric, or repeated, so they can interact with multimeric proteins—those composed of more than one subunit. Most DNA-binding proteins are dimers (some are even tetramers), and this greatly enhances the binding between DNA and protein because the two protein subunits

bind cooperatively. Having one at the binding site automatically increases the concentration of the other. This boost in concentration is important because DNA-binding proteins are generally present in the cell in very small quantities.

Another way of looking at the advantage of dimeric DNA-binding proteins uses the concept of *entropy*. Entropy can be considered a measure of disorder in the universe. It probably does not come as a surprise to you to learn that entropy, or disorder, naturally tends to increase with time. Think of what happens to the disorder of your room, for example. The disorder increases with time until you expend energy to straighten it up. Thus, it takes energy to push things in the opposite of the natural direction—to create order out of disorder, or make the entropy of a system decrease.

A DNA–protein complex is more ordered than the same DNA and protein independent of each other, so bringing them together causes a decrease in entropy. Binding *two* protein subunits, independently of each other, causes twice the decrease in entropy. But if the two protein subunits are already stuck together in a dimer, orienting one relative to the DNA automatically orients the other, so the entropy change is much less than in independent binding, and therefore requires less energy. Looking at it from the standpoint of the DNA–protein complex, releasing the dimer from the DNA does not provide the same entropy gain as releasing two independently bound proteins would, so the protein and DNA stick together more tightly.

SUMMARY Multimeric DNA-binding proteins have an inherently higher affinity for binding sites on DNA than do multiple monomeric proteins that bind independently of one another.

9.4 DNA-Binding Proteins: Action at a Distance

So far, we have dealt primarily with DNA-binding proteins that govern events that occur very nearby. For example, the *lac* repressor bound to its operator interferes with the activity of RNA polymerase at an adjacent DNA site; or λ repressor stimulates RNA polymerase binding at an adjacent site. However, numerous examples exist in which DNA-binding proteins can influence interactions at remote sites in the DNA. We will see that this phenomenon is common in eukaryotes, but several prokaryotic examples occur as well.

The *gal* Operon

In 1983, S. Adhya and colleagues reported the unexpected finding that the *E. coli gal* operon, which codes for enzymes needed to metabolize the sugar galactose, has two distinct operators, about 97 bp apart. One is located where you would expect to find an operator, adjacent to the *gal* promoter. This one is called O_E, for "external" operator. The other is called O_I, for "internal" operator and is located within the first structural gene, *galE*. The downstream operator was discovered by genetic means: O^c mutations were found that mapped to the *galE* gene instead of to O_E. One way to explain the function of two separated operators is by assuming that they both bind to repressors, and the repressors interact by **looping out** the intervening DNA, as pictured in Figure 9.15. We have already seen examples of this kind of repression by looping out in our discussion of the *lac* and *ara* operons in Chapter 7.

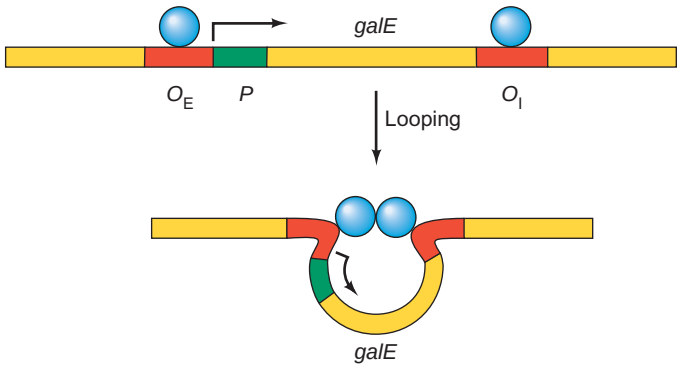


Figure 9.15 Repression of the *gal* operon. The *gal* operon has two operators (red): one external (O_E), adjacent to the promoter (green), and one internal (O_I), within the *galE* gene (yellow). Repressor molecules (blue) bind to both operators and appear to interact by looping out the intervening DNA (bottom).

Duplicated λ Operators

The brief discussion of the *gal* operon just presented strongly suggests that proteins interact over a distance of almost 100 bp, but provided no direct evidence for this contention. Ptashne and colleagues used an artificial system to obtain such evidence. The system was the familiar λ operator–repressor combination, but it was artificial in that the experimenters took the normally adjacent operators and separated them to varying extents. We have seen that repressor dimers normally bind cooperatively to O_R1 and O_R2 when these operators are adjacent. The question is this: Do repressor dimers still bind cooperatively to the operators when they are separated? The answer is that they do, as long as the operators lie on the same face of the DNA double helix. This finding supports the hypothesis that repressors bound to separated *gal* operators probably interact by DNA looping.

Ptashne and coworkers used two lines of evidence to show cooperative binding to the separated λ promoters: DNase footprinting and electron microscopy. If we

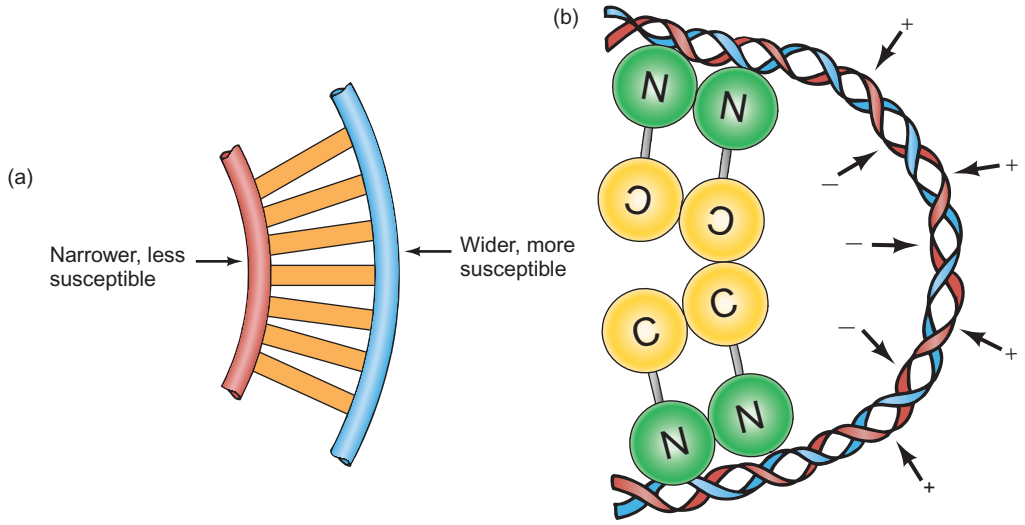


Figure 9.16 Effect of DNA looping on DNase susceptibility. **(a)** Simplified schematic diagram. The double helix is depicted as a railroad track to simplify the picture. The backbones are in red and blue, and the base pairs are in orange. As the DNA bends, the strand on the inside of the bend is compressed, restricting access to DNase. By the same token, the strand on the outside is stretched, making it easier for DNase to attack. **(b)** In a real helix each strand alternates being on the inside and the outside of the bend. Here, two dimers of a DNA-binding protein (λ repressor in this example) are interacting at separated sites, looping out the DNA in between. This stretches the DNA on the outside of the loop, opening it up to DNase I attack (indicated by + signs). Conversely, looping compresses the DNA on the inside of the loop, obstructing access to DNase I (indicated by the – signs). The result is an alternating pattern of higher and lower sensitivity to DNase in the looped region. Only one strand (red) is considered here, but the same argument applies to the other. (*Source:* (*b*) Adapted from Hochschild A. and M. Ptashne, Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell* 44:685, 1986.)

DNase-footprint two proteins that bind independently to remote DNA sites, we see two separate footprints. However, if we footprint two proteins that bind cooperatively to remote DNA sites through DNA looping, we see two separate footprints just as in the previous example, but this time we also see something interesting in between that does not occur when the proteins bind independently. This extra feature is a repeating pattern of insensitivity, then hypersensitivity to DNase. The reason for this pattern is explained in Figure 9.16. When the DNA loops out, the bend in the DNA compresses the base pairs on the inside of the loop, so they are relatively protected from DNase. On the other hand, the base pairs on the outside of the loop are spread apart more than normal, so they become extra sensitive to DNase. This pattern repeats over and over as we go around and around the double helix.

Using this assay for cooperativity, Ptashne and colleagues performed DNase footprinting on repressor bound to DNAs in which the two operators were separated by an integral or nonintegral number of double-helical turns. Figure 9.17a shows an example of cooperative binding, when the two operators were separated by 63 bp—almost exactly six double-helical turns. We can see the repeating pattern of lower and higher DNase sensitivity in between the two binding sites. By contrast, Figure 9.17b presents an example of noncooperative binding, in which the two operators were separated by 58 bp—just 5.5 double-helical turns. Here we see no evidence of a repeating pattern of DNase sensitivity between the two binding sites.

Electron microscopy experiments enabled Ptashne and coworkers to look directly at repressor–operator complexes with integral and nonintegral numbers of double-helical turns between the operators to see if the DNA in the former case really loops out. As Figure 9.18 shows, it does loop out. It is clear when such looping out is occurring, because the DNA is drastically bent. By contrast, Ptashne and colleagues almost never observed bent DNA when the two operators were separated by a nonintegral number of double-helical turns. Thus, as expected, these DNAs have a hard time looping out. These experiments demonstrate clearly that proteins binding to DNA sites separated by an integral number of double-helical turns can bind cooperatively by looping out the DNA in between.

SUMMARY When λ operators are separated by an integral number of double-helical turns, the DNA in between can loop out to allow cooperative binding. When the operators are separated by a nonintegral number of double-helical turns, the proteins have to bind to opposite faces of the DNA double helix, so no cooperative binding can take place.

Enhancers

Enhancers are nonpromoter DNA elements that bind protein factors and stimulate transcription. By definition, they

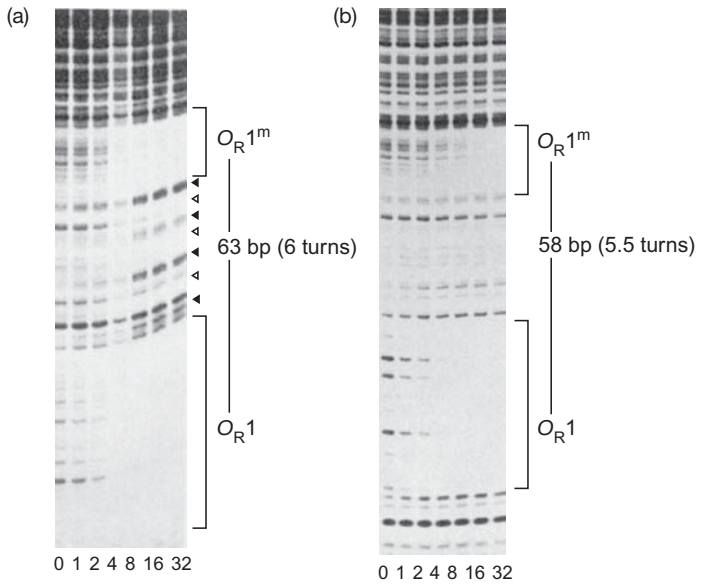


Figure 9.17 DNase footprints of dual operator sites. **(a)** Cooperative binding. The operators are almost exactly six double-helical turns apart (63 bp), and an alternating pattern of enhanced and reduced cleavage by DNase I appears between the two footprints when increasing amounts of repressor are added. The enhanced cleavage sites are denoted by filled arrowheads, the reduced cleavage sites by open arrowheads. This suggests looping of DNA between the two operators on repressor binding. **(b)** Noncooperative binding. The operators are separated by a nonintegral number of double-helical turns (58 bp, or 5.5 turns). No alternating pattern of DNase susceptibility appears on repressor binding, so the repressors bind at the two operators independently, without DNA looping. In both (a) and (b), the number at the bottom of each lane gives the amount of repressor monomer added, where 1 corresponds to 13.5 nM repressor monomer in the assay, 2 corresponds to 27 nM repressor monomer, and so on. (*Source:* Adapted from Hochschild, A. and M. Ptashne, Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell* 44 (14 Mar 1986) f. 3a&4, p. 683.)

can act at a distance. Such elements have been recognized in eukaryotes since 1981, and we will discuss them at length in Chapter 12. More recently, enhancers have also been found in prokaryotes. In 1989, Popham and coworkers described an enhancer that aids in the transcription of genes recognized by an auxiliary σ-factor in *E. coli:* σ^{54}. We encountered this factor in Chapter 8; it is the σ-factor, also known as σ^N, that comes into play under nitrogen starvation conditions to transcribe the *glnA* gene from an alternative promoter.

The σ^{54} factor is defective. DNase footprinting experiments demonstrate that it can cause the $E\sigma^{54}$ holoenzyme to bind stably to the *glnA* promoter, but it cannot do one of the important things normal σ-factors do: direct the formation of an open promoter complex. Popham and coworkers assayed this function in two ways: heparin resistance and DNA methylation. When polymerase forms an open promoter complex, it is bound very tightly to DNA. Adding heparin as a DNA competitor does not inhibit the polymerase. On the other hand, when polymerase forms a closed promoter complex, it is relatively loosely bound and will dissociate at a much higher rate. Thus, it is subject to inhibition by an excess of the competitor heparin. Furthermore, when polymerase forms an open promoter complex, it exposes the cytosines in the melted DNA to methylation by DMS. Because no melting occurs in the closed promoter complex, no methylation takes place.

By both these criteria—heparin sensitivity and resistance to methylation—$E\sigma^{54}$ fails to form an open promoter complex. Instead, another protein, NtrC (the product of the *ntrC* gene), binds to the enhancer and helps $E\sigma^{54}$ form an open promoter complex. The energy for the DNA melting comes from the hydrolysis of ATP, performed by an ATPase domain of NtrC.

How does the enhancer interact with the promoter? The evidence strongly suggests that DNA looping is involved. One clue is that the enhancer has to be at least 70 bp away from the promoter to perform its function. This would allow enough room for the DNA between the promoter and enhancer to loop out. Moreover, the enhancer can still function even if it and the promoter are on separate DNA molecules, as long as the two molecules are linked in a catenane, as shown in Figure 9.19. This would still allow the enhancer and promoter to interact as they would during looping, but it precludes any mechanism (e.g., altering the degree of supercoiling or sliding proteins along the DNA) that requires the two elements to be on the same DNA molecule. We will discuss this phenomenon in more detail in Chapter 12. Finally, and perhaps most tellingly, we can actually observe the predicted DNA loops between NtrC bound to the enhancer and the σ^{54} holoenzyme bound to the promoter. Figure 9.20 shows the results of electron microscopy experiments performed by Sydney Kustu, Harrison Echols, and colleagues with cloned DNA containing the enhancer–*glnA* region. These workers inserted 350 bp of DNA between the enhancer and promoter to make the loops easier to see. The polymerase holoenzyme stains more darkly than NtrC in most of these electron micrographs, so we can distinguish the two proteins at the bases of the loops, just as we would predict if the two proteins interact by looping out the DNA in between. The loops were just the right size to account for the length of DNA between the enhancer and promoter.

Phage T4 provides an example of an unusual, mobile enhancer that is not defined by a set base sequence. Transcription of the late genes of T4 depends on DNA replication; no late transcription occurs until the phage DNA begins to replicate. One reason for this linkage between late transcription and DNA replication is that the late phage σ-factor (σ^{55}), like σ^{54} of *E. coli,* is defective. It cannot function without an enhancer. But the late T4 enhancer is not a fixed DNA sequence like the NtrC-binding site. Instead, it is the DNA replicating fork. The enhancer-binding protein, encoded by phage genes 44, 45, and 62, is part of the phage DNA replicating machinery. Thus, this protein migrates along with the

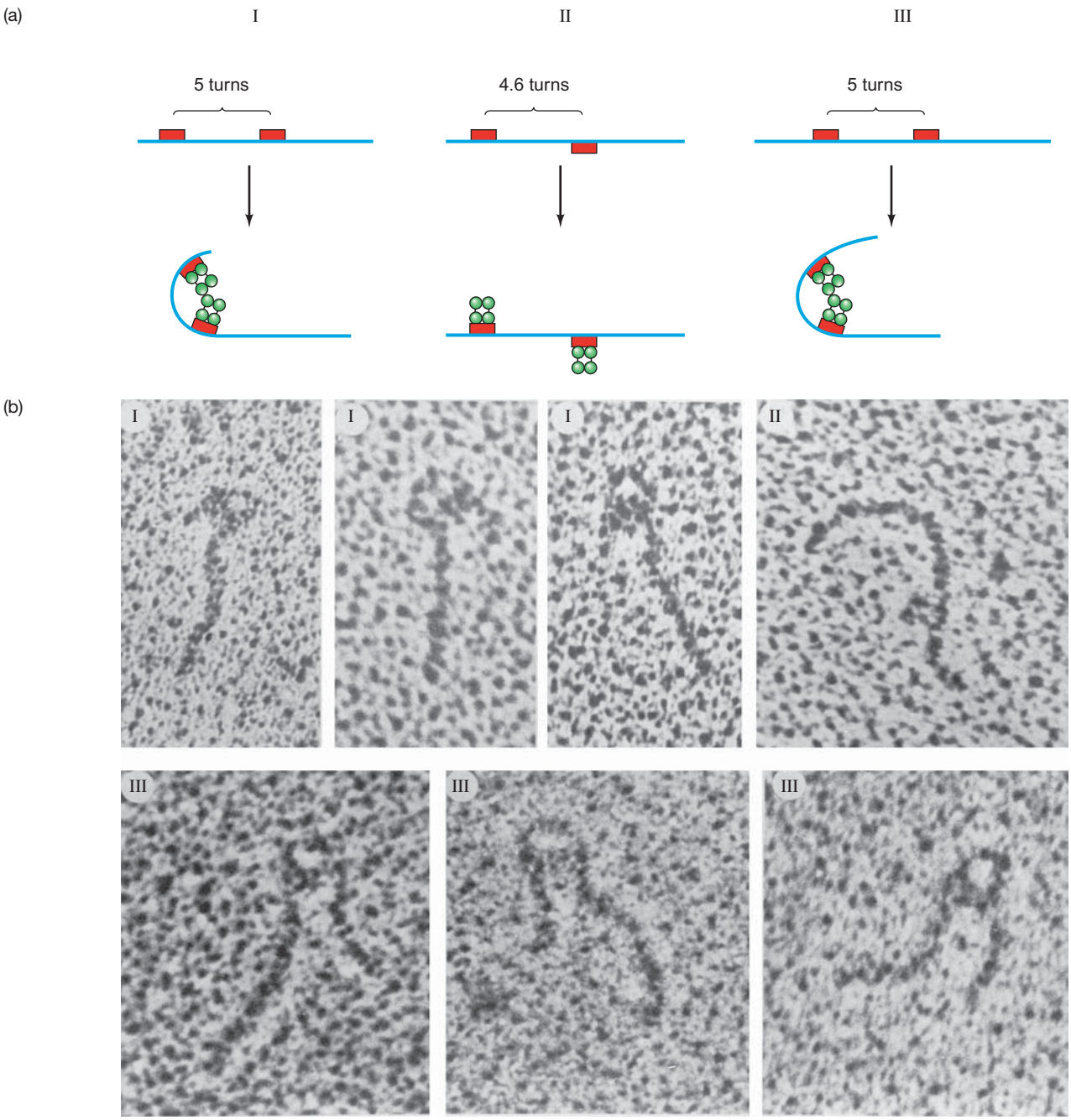


Figure 9.18 Electron microscopy of λ repressor bound to dual operators. (a) Arrangement of dual operators in three DNA molecules. In I, the two operators are five helical turns apart near the end of the DNA; in II, they are 4.6 turns apart near the end; and in III they are five turns apart near the middle. The arrows in each case point to a diagram of the expected shape of the loop due to cooperative binding of repressor to the two operators. In II, no loop should form because the two operators are not separated by an integral number of helical turns and are consequently on opposite sides of the DNA duplex. **(b)** Electron micrographs of the protein–DNA complexes. The DNA types [I, II, or III from panel (a) used in the complexes are given at the upper left of each picture. The complexes really do have the shapes predicted in panel (a). (*Source:* (a) Griffith et al., DNA loops induced by cooperative binding of lambda repressor. *Nature* 322 (21 Aug 1986) f. 2, p. 751. © Macmillan Magazines Ltd.)

replicating fork, which keeps it in contact with the moving enhancer.

One can mimic the replicating fork in vitro with a simple nick in the DNA, but the polarity of the nick is important: It works as an enhancer only if it is in the nontemplate strand. This suggests that the T4 late enhancer probably does not act by DNA looping because polarity does not matter in looping. Furthermore, unlike typical enhancers

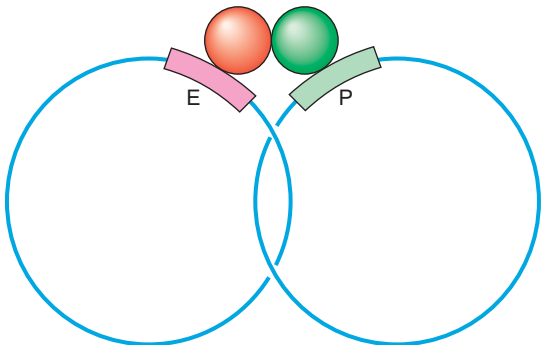


Figure 9.19 Interaction between two sites on separate but linked DNA molecules. An enhancer (E, pink) and a promoter (P, light green) lie on two separate DNA molecules that are topologically linked in a catenane (intertwined circles). Thus, even though the circles are distinct, the enhancer and promoter cannot ever be far apart, so interactions between proteins that bind to them (red and green, respectively) are facilitated.

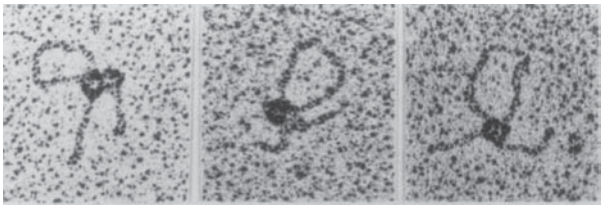


Figure 9.20 Looping the *glnA* promoter–enhancer region. Kustu, Echols, and colleagues moved the *glnA* promoter and enhancer apart by inserting a 350-bp DNA segment between them, then allowed the NtrC protein to bind to the enhancer, and RNA polymerase to bind to the promoter. When the two proteins interacted, they looped out the DNA in between, as shown in these electron micrographs. (*Source:* Su, W., S. Porter, S. Kustu, and H. Echols, DNA-looping and enhancer activity: Association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proceedings of the National Academy of Sciences USA 87* (July 1990) f. 4, p. 5507.)

such as the *glnA* enhancer, the T4 late enhancer must be on the same DNA molecule as the promoters it controls. It does not function *in trans* as part of a catenane. This argues against a looping mechanism.

SUMMARY The *E. coli glnA* gene is an example of a prokaryotic gene that depends on an enhancer for its transcription. The enhancer binds the NtrC protein, which interacts with polymerase bound to the promoter at least 70 bp away. Hydrolysis of ATP by NtrC allows the formation of an open promoter complex so transcription can take place. The two proteins appear to interact by looping out the DNA in between. The phage T4 late enhancer is mobile; it is part of the phage DNA-replication apparatus. Because this enhancer must be on the same DNA molecule as the late promoters, it probably does not act by DNA looping.

SUMMARY

The repressors of the λ-like phages have recognition helices that fit sideways into the major groove of the operator DNA. Certain amino acids on the DNA side of the recognition helix make specific contact with bases in the operator, and these contacts determine the specificity of the protein–DNA interactions. Changing these amino acids can change the specificity of the repressor. The λ repressor and Cro protein share affinity for the same operators, but they have microspecificities for O_R1 or O_R3, determined by interactions between different amino acids in the recognition helices of the two proteins and base pairs in the different operators.

The cocrystal structure of a λ repressor fragment with an operator fragment shows many details about how the protein and DNA interact. The most important contacts occur in the major groove, where amino acids on the recognition helix, and other amino acids, make hydrogen bonds with the edges of DNA bases and with the DNA backbone. Some of these hydrogen bonds are stabilized by hydrogen bond networks involving two amino acids and two or more sites on the DNA. The structure derived from the cocrystal is in almost complete agreement with previous biochemical and genetic data.

X-ray crystallography of a phage 434 repressor-fragment/operator-fragment complex shows probable hydrogen bonding between amino acid residues in the recognition helix and base pairs in the repressor. It also reveals a potential van der Waals contact between an amino acid in the recognition helix and a base in the operator. The DNA in the complex deviates significantly from its normal regular shape. It bends somewhat to accommodate the necessary base/amino acid contacts. Moreover, the central part of the helix, between the two half-sites, is wound extra tightly, and the outer parts are wound more loosely than normal. The base sequence of the operator facilitates these departures from normal DNA shape.

The *trp* repressor requires tryptophan to force the recognition helices of the repressor dimer into the proper position for interacting with the *trp* operator.

A DNA-binding protein can interact with the major or minor groove of the DNA (or both). The four different base pairs present four different hydrogen-bonding profiles to amino acids approaching either the major or minor DNA groove, so a DNA-binding protein can recognize base pairs in the DNA even though the two strands do not separate.

Multimeric DNA-binding proteins have an inherently higher affinity for binding sites on DNA than do multiple monomeric proteins that bind independently of one another. The advantage of multimeric proteins is that they can bind cooperatively to DNA.

When λ operators are separated by an integral number of helical turns, the DNA in between can loop out to allow cooperative binding. When the operators are separated by a nonintegral number of helical turns, the proteins have to bind to opposite faces of the DNA double helix, so no cooperative binding can take place.

The *E. coli glnA* gene is an example of a bacterial gene that depends on an enhancer for its transcription. The enhancer binds the NtrC protein, which interacts with polymerase bound to the promoter at least 70 bp away. Hydrolysis of ATP by NtrC allows the formation of an open promoter complex so transcription can take place. The two proteins appear to interact by looping out the DNA in between. The phage T4 late enhancer is mobile; it is part of the phage DNA-replication apparatus. Because this enhancer must be on the same DNA molecule as the late promoters, it probably does not act by DNA looping.

REVIEW QUESTIONS

1. Draw a rough diagram of a helix-turn-helix domain interacting with a DNA double helix.
2. Describe and give the results of an experiment that shows which amino acids are important in binding between λ-like phage repressors and their operators. Present two methods of assaying the binding between the repressors and operators.
3. In general terms, what accounts for the different preferences of λ repressor and Cro for the three operator sites?
4. Glutamine and asparagine side chains tend to make what kind of bonds with DNA?
5. Methylene and methyl groups on amino acids tend to make what kind of bonds with DNA?
6. What is meant by the term *hydrogen bond network* in the context of protein–DNA interactions?
7. Draw a rough diagram of the "reading head" model to show the difference in position of the recognition helix of the *trp* repressor and aporepressor, with respect to the *trp* operator.
8. Draw a rough diagram of the "salami sandwich" model to explain how adding tryptophan to the *trp* aporepressor causes a shift in conformation of the protein.
9. In one sentence, contrast the orientations of the λ and *trp* repressors relative to their respective operators.
10. Explain the fact that protein oligomers (dimers or tetramers) bind more successfully to DNA than monomeric proteins do.
11. Use a diagram to explain the alternating pattern of resistance and elevated sensitivity to DNase in the DNA between two separated binding sites when two proteins bind cooperatively to these sites.
12. Describe and give the results of a DNase footprinting experiment that shows that λ repressor dimers bind cooperatively to two operators separated by an integral number of DNA double-helical turns, but noncooperatively to two operators separated by a nonintegral number of turns.
13. Describe and give the results of an electron microscopy experiment that shows the same thing as the experiment in the preceding question.
14. In what way is σ^{54} defective?
15. What substances supply the missing function to σ^{54}?
16. Describe and give the results of an experiment that shows that DNA looping is involved in the enhancement of the *E. coli glnA* locus.
17. In what ways is the enhancer for phage T4 σ^{55} different from the enhancer for the *E. coli* σ^{54}?

ANALYTICAL QUESTIONS

1. An asparagine in a DNA-binding protein makes an important hydrogen bond with a cytosine in the DNA. Changing this glutamine to alanine prevents formation of this hydrogen bond and blocks the DNA–protein interaction. Changing the cytosine to thymine restores binding to the mutant protein. Present a plausible hypothesis to explain these findings.
2. You have the following working hypothesis: To bind well to a DNA-binding protein, a DNA target site must twist less tightly and widen the narrow groove between base pairs 4 and 5. Suggest an experiment to test your hypothesis.
3. Draw a T–A base pair. Based on that structure, draw a line diagram indicating the relative positions of the hydrogen bond acceptor and donor groups in the major and minor grooves. Represent the horizontal axis of the base pair by two segments of a horizontal line, and the relative horizontal positions of the hydrogen bond donors and acceptors by vertical lines. Let the lengths of the vertical lines indicate the relative vertical positions of the acceptors and donors. What relevance does this diagram have for a protein that interacts with this base pair?

SUGGESTED READINGS

General References and Reviews

Geiduschek, E.P. 1997. Paths to activation of transcription. *Science* 275:1614–16.

Kustu, S., A.K. North, and D.S. Weiss. 1991. Prokaryotic transcriptional enhancers and enhancer-binding proteins. *Trends in Biochemical Sciences* 16:397–402.

Schleif, R. 1988. DNA binding by proteins. *Science* 241:1182–87.

Research Articles

Aggarwal, A.K., D.W. Rodgers, M. Drottar, M. Ptashne, and S.C. Harrison. 1988. Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. *Science* 242:899–907.

Griffith, J., A. Hochschild, and M. Ptashne. 1986. DNA loops induced by cooperative binding of λ repressor. *Nature* 322:750–52.

Herendeen, D.R., G.A. Kassavetis, J. Barry, B.M. Alberts, and E.P. Geiduschek. 1990. Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. *Science* 245:952–58.

Hochschild, A., J. Douhann III, and M. Ptashne. 1986. How λ repressor and λ cro distinguish between O_R1 and O_R3. *Cell* 47:807–16.

Hochschild, A. and M. Ptashne. 1986. Cooperative binding of λ repressors to sites separated by integral turns of the DNA helix. *Cell* 44:681–87.

Jordan, S.R. and C.O. Pabo. 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor–operator interactions. *Science* 242:893–99.

Popham, D.L., D. Szeto, J. Keener, and S. Kustu. 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* 243:629–35.

Sauer, R.T., R.R. Yocum, R.F. Doolittle, M. Lewis, and C.O. Pabo. 1982. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. *Nature* 298:447–51.

Schevitz, R.W., Z. Otwinowski, A. Joachimiak, C.L. Lawson, and P. B. Sigler. 1985. The three-dimensional structure of *trp* repressor. *Nature* 317:782–86.

Su, W., S. Porter, S. Kustu, and H. Echols. 1990. DNA looping and enhancer activity: Association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proceedings of the National Academy of Sciences USA* 87:5504–8.

Wharton, R.P. and M. Ptashne. 1985. Changing the binding specificity of a repressor by redesigning an α-helix. *Nature* 316:601–5.

Zhang, R.-g., A. Joachimiak, C.L. Lawson, R.W. Schevitz, Z. Otwinowski, and P.B. Sigler. 1987. The crystal structure of *trp* aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature* 327:591–97.