Structural Basis of Protein–Nucleic Acid Interactions

DAVID L. OLLIS*

Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, Illinois 60201

STEPHEN W. WHITE

Department of Biology, Brookhaven National Laboratory, Upton, New York 11973

Received March 9, 1987 (Revised Manuscript Received June 12, 1987)

Contents

I. Introduction

Interactions between protein and nucleic acid molecules are central to many of the vital processes in molecular biology.[1] Research into this area has been enriched in the last few years by the elucidation of the structures of a number of DNA binding proteins and, more importantly, the structures of complexes formed between proteins and DNA. The structural data base of these types of interactions is now significant and worth reviewing. Reviews covering other aspects of the interactions of proteins and nucleic acids can be found elsewhere.[2-4]

The importance of structural data should not be underestimated. Direct visualization of complexes at the molecular level provides the only unambiguous method of understanding the nature of these interactions. Unfortunately, the crystal structures of only three such complexes have been determined, but these have provided invaluable information and will therefore be described in detail in this review.

Many nucleic acid binding proteins have been solved in the uncomplexed form, and these have allowed the formulation of models for the interaction. In most cases, the models are also based on independent genetic and biochemical data and suggest further experiments to test them. These data, which will be described briefly where appropriate, are important as they sometimes lead to radical changes in the proposed model. It should be emphasized that models for protein–nucleic acid interaction based on the structure of the protein alone suffer from a number of limitations. Structural changes may occur in the protein and/or the nucleic acid at the time of binding, and the detailed nature of the interactions can only be guessed at.

For the purposes of this review, the term nucleic acids is meant to include only oligomeric molecules, and reference will not be made to the way in which proteins interact with molecules such as ATP. This has been described elsewhere.[2] This review draws mainly upon well-defined crystal structure data. Only those proteins whose mode of interaction with nucleic acid is reasonably well determined have been included. It will be apparent that the available protein structures carry out a variety of functions and, as will be explained, interact with nucleic acids in a variety of ways. It should be noted, however, that this review covers only a small fraction of this class of protein and of necessity represents mainly those proteins that have proved amenable to crystallization. The enzymes responsible for transcription and recombination are good examples of important enzymes for which structures are not available.

II. General Considerations

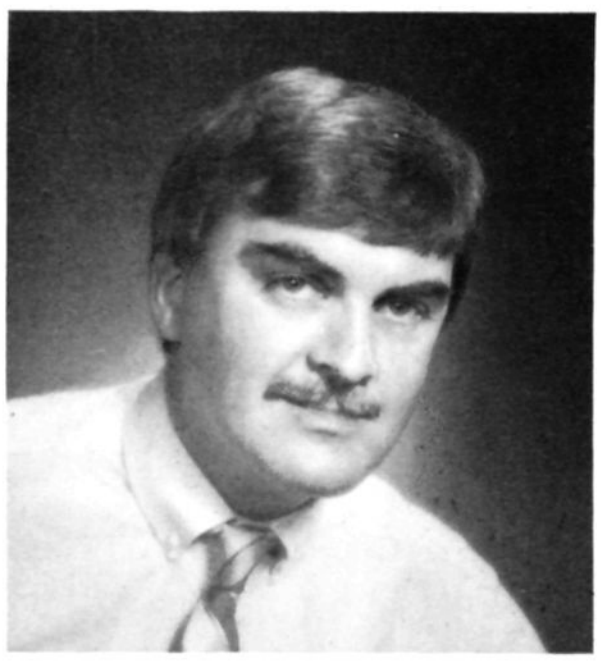
At this point it is worth making a few statements about nucleic acid structure and describing some of the early ideas on the ways in which they might interact with proteins.

A. Nucleic Acid Structure

The structures of DNA and RNA have been fully described elsewhere,[2] but a brief description emphasizing the features that are important to this review is useful. Nucleic acids are polymers in which the repeat unit is a nucleotide consisting of a phosphate group, a 5-carbon sugar moiety, and a purine or pyrimidine base. Ribonucleic acid (RNA) contains the sugar ribose, and the bases adenine (A), uracil (U), guanine (G), and cytosine (C). Deoxyribonucleic acid (DNA) contains the sugar deoxyribose and the same bases as RNA apart from thymine (T), which replaces uracil. The nucleotides are linked by phosphodiester bonds, creating a

0009-2665/87/0787-0981$06.50/0

David Ollis studied chemistry at the University of New South Wales for his undergraduate degree. His graduate degree involved structural studies of heme proteins and was carried out in the Department of Inorganic Chemistry at Sydney University. During a postdoctoral period in the Department of Molecular Biophysics and Biochemistry at Yale University, he was occupied principally by the structure determination of the large fragment of DNA Polymerase I. At present he has appointments in the Department of Biochemistry, Molecular Biology, and Cell Biology and the Department of Chemistry at Northwestern University in Evanston, IL.



Stephen White was born in London, England, and obtained a B.Sc. in Biochemistry from Bristol University and a D. Phil. in Molecular Biophysics from Oxford University. He remained at Oxford for 2 years as a Junior Research Fellow at Wolfson College, and then moved to the Max Planck Institut fuer Molekulare Genetik, West Berlin, where he studied the structures of bacterial ribosomal proteins and histone-like proteins. He is currently an Associate Biophysicist in the Department of Biology, Brookhaven National Laboratory, New York, and is investigating protein structures by use of X-ray and neutron diffraction techniques.

sugar phosphate backbone via the 3- and 5-positions of the sugar. The base is linked to the 1-position of the sugar. The phosphodiester is a strong acid, and the polymer has 1 negative charge/repeat unit. Most DNAs are double-stranded molecules in which the two strands run in opposite directions and are connected by hydrogen-bond interactions between adenine and thymine, and guanine and cytosine. These two so-called base pairs are almost planar and virtually identical in terms of their geometry and stack on top of each other in the center of the molecule to create the double-helical structure first proposed by Watson and Crick.[5] RNA molecules are mostly single stranded but can contain a considerable amount of double-helical structure formed by intra-strand base pair interactions. The resulting loops and hairpins are often referred to as RNA secondary structure. Nucleic acids vary enormously in size from the small transfer (t) RNA molecules (75–85 residues) to the giant DNA molecules of some 10^8 residues, which are found in the genome of higher animals.

Nucleic acids are able to adopt several types of helical conformations. The most common in DNA is B-DNA, which exists under physiological conditions and is therefore the probable dominant form in vivo. The important helical parameters are 10 residues/turn, a rise per residue of ~3.4 Å, and a diameter of ~22 Å. The planar bases are perpendicular to the helical axis, and the molecule contains two distinct grooves—a large major groove and a smaller minor groove—on opposite sides of the helix. B-DNA, like all the helical structures, has two symmetry axes perpendicular to the helix axis. The first is a dyad axis at the level of each base pair. This is a true dyad with respect to the sugar phosphate backbone, but not with respect to the complete molecule. The second is a C_2 axis between each base pair. This is a true axis in regions of DNA that contain palindromic sequences and is an important feature in several protein–DNA interactions. The other major helical form is A form, which is adopted by RNA duplexes, RNA–DNA hybrids, and DNA at low relative humidity. This form has 11 residues/turn, with a larger diameter than the B form, and the base pairs are tilted by 20° with respect to the helix axis. These helical parameters were originally determined from fiber diffraction patterns and represent an average over the complete molecule. Recently, the high-resolution structures of several self-complementary oligonucleotides were determined.[6,7] These basically confirm the helical parameters but also show that the local nucleotide sequence can cause considerable variations[8,9] and induce a bending of the DNA.[10]

B. Types of Interaction

These general features of nucleic acid structures allow one to predict the types of interaction that are likely to form the basis of their interactions with proteins:[2] a. Salt bridges between the negatively charged phosphate groups and positively charged side chains on the protein. Such interactions are ideal for nonspecific binding, since the sugar phosphate backbone is a common feature of all nucleic acids. b. Hydrogen bonding between the various acceptor and donor groups on the molecules. This has been predicted and found to be the principal basis of specific protein–DNA interactions. c. Aromatic ring stacking interactions between the bases and the aromatic protein side chains. d. Hydrophobic interactions between the bases and the nonpolar protein side chains.

There have also been predictions concerning the type of protein secondary structure or motif that might bind to nucleic acids, in particular, DNA. Protein α helixes and β pleated sheet structures have rather specific geometrical properties and dimensions,[11] some of which are complementary to those of the DNA helix. Zubay

and Doty[12] noted that the α helix would fit snugly into the B-DNA major groove. Others suggested that an antiparallel β pleated sheet has the correct twist and helical structure for interaction with duplex (double-stranded) RNA[13] and DNA.[14] It was shown by model building that an antiparallel β ribbon (equivalent to a two-strand sheet) would fit into the B-DNA minor groove.[14] Not only are the relative dimensions comparable, but also the antiparallel nature of the two structures.

C. Early Models

It is not difficult to envisage how a protein might bind nonspecifically to DNA through interactions with the sugar phosphate backbone. A more interesting problem, and one that is central to many important processes in molecular biology, concerns the recognition by a protein of a specific DNA sequence. This, of necessity, requires that a protein "reads" the sequence of bases along the DNA/RNA molecule. Many people suggested that hydrogen bonds have the ideal properties for this role. They have both high specificity and rather rigid geometrical requirements for their formation. Thus, a constellation of hydrogen-bond acceptors and donors formed by a certain sequence of bases in the DNA could be recognized by a complementary constellation on the protein. Seeman et al.[15] analyzed the hydrogen-bond donors and acceptors that each base pair exposes to the major and minor grooves of DNA. They concluded that more than one hydrogen bond is necessary to uniquely identify a particular base pair and that discrimination is more easily achieved via the major groove, which exposes more acceptor and donor groups. They also proposed that a potentially powerful discrimination would result from the use of one protein side chain to hydrogen bond to two functional groups. This proposal was partly based on hydrogen-bonding patterns in tRNA structures.

Von Hippel and Berg[16] undertook a more theoretical examination of specific DNA–protein interactions. They pointed out that the stability gained by the correct hydrogen bonding in a DNA–protein complex is much less than the instability resulting from incorrect hydrogen-bond complementarities. This is because free DNA and protein can hydrogen bond to water, but unmatched donors and acceptors in a wrong complex may be buried and unable to do so. They also concluded that the relatively weak specific interactions would need to be supplemented by nonspecific electrostatic interactions to achieve the observed binding affinities. They proposed that the protein has an alternative conformation that allows totally nonspecific binding to DNA until the correct specific binding site has been located. This would avoid the destabilizing effects of hydrogen-bond mismatch and allow the protein to slide along the molecule in a one-dimensional diffusion process to facilitate the location of the target site.

III. Nonspecific Protein–Nucleic Acid Interactions

A. Gene 5 from Bacteriophage fd

Bacteriophage fd is a rodlike structure containing a single-stranded circle of DNA covered by multiple copies of a coat protein. The DNA of 6408 bases has been completely sequenced and contains 10 genes. Gene 5 codes for a small protein of molecular weight 9700 Da (87 amino acids), which functions during replication to bind viral daughter strands and prevent further DNA synthesis and attack from nucleases. The complexes contain some 1300 gene 5 proteins/6408 DNA bases and form helical rods. The protein is dimeric, and its strong preference for single-stranded DNA gives it potent DNA helix-destabilizing properties.[17]

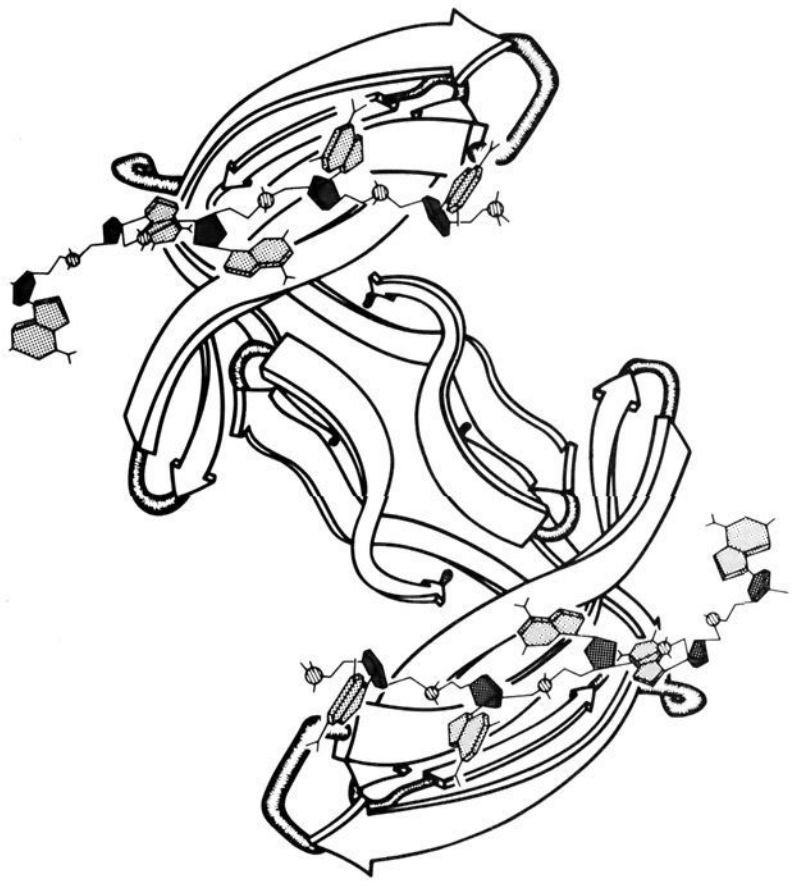


Figure 1. Stylized drawing of the single-stranded binding protein from phage fd and the paths of the two strands of the DNA. The interaction was deduced from a combination of model building and biochemical data. Reprinted with permission from ref 19. Copyright 1984 American Chemical Society.

The structure of gene 5 protein has been refined to high (2.3 Å) resolution,[18] and it is primarily a β sheet structure with several interconnecting loops (see Figure 1). The fold can be described as three β loops extending from a common hydrophobic core. The loops are referred to as the DNA binding loop, the complex loop, and the dyad loop. Dimerization is mediated by the dyad and complex loops and involves considerable interaction between the monomers. Since each monomer is a rather open structure, it is likely to be unstable in the absence of a dimeric partner. In the dimer, the dyad loops and the N-terminal β strands form a central six-stranded β barrel, which is perpendicular to and bisected by the intermolecular 2-fold axis. Dimer stability derives from approximately 10 ionic interactions and the hydrophobic interior of the β barrel, which is contiguous with the cores of the distal regions of each monomer.

There are many independent spectral and chemical data that show which amino acids are important for the interaction of this protein with DNA (reviewed in ref 17 and 19). NMR data indicate that a number of aromatic residues are involved in base stacking interactions and also that lysine and arginine residues interact with the phosphate backbone. These spectroscopic studies were confirmed by chemical modification and cross-

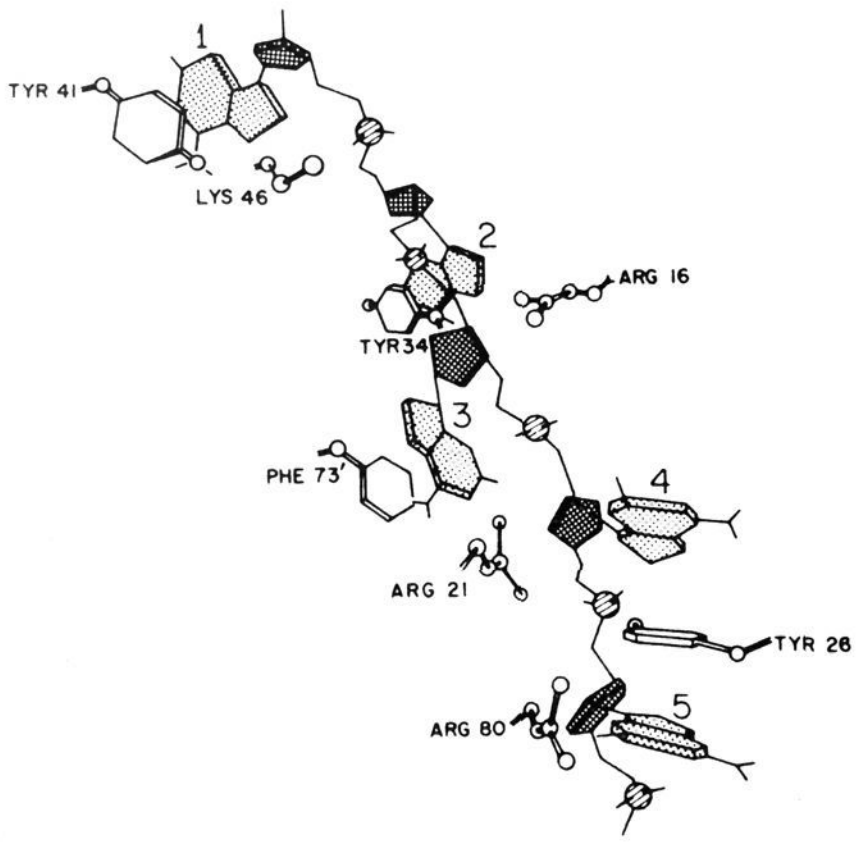


Figure 2. Closer view of the interactions between the gene 5 protein from phage fd and one strand of DNA. The interactions were deduced from a combination of model building and biochemical data. See text for details. Reprinted with permission from ref 19. Copyright 1984 American Chemical Society.

linking studies. All the residues identified by the above studies are in, or close to, a shallow channel on the protein surface. The molecular 2-fold creates two such channels with reverse polarity and separated by approximately 30 Å (see Figure 1). Therefore, they are ideally disposed to bind and keep apart the antiparallel strands of a DNA molecule. For the most part, each channel is confined to one monomer. The only cross-interaction occurs through the β bend at the end of the dyad loop. The length of each channel is consistent with five interacting nucleotides, if it is assumed that the DNA is in an extended conformation. This number agrees with other estimates.

A detailed model has been proposed for the protein–DNA complex[19] that features ionic interactions between the DNA phosphates and the lysine and arginine side chains and aromatic ring stacking interactions between the bases and tyrosine and phenylalanine side chains (see Figure 2). This model agrees well with most of the independent data outlined above apart from the role of two tyrosine residues. One of these residues is apparently not accessible to chemical modification, although it is clearly exposed in the structure, while the opposite appears to be true of the other tyrosine. The model required some alterations to the refined structure, in particular the orientation of some surface side chains and a movement of the DNA binding loop. This is not unreasonable, since these parts of the structure exhibit high temperature factors[18] and are therefore rather flexible. The model, in fact, predicts how the movement of the DNA binding loop might be triggered by the binding of the DNA.

There are several other proteins from a variety of sources that also bind cooperatively and nonspecifically to single-stranded DNA. Examples include SSB (single-stranded binding) protein from *Escherichia coli*,[20] proteins encoded by bacteriophages T4 and T7,[20] and the Rec A protein also from *E. coli*.[21] The mechanism whereby the fd gene 5 protein binds to single-stranded DNA, namely exposed aromatic residues that interact with the bases and basic residues that interact with the phosphate backbone of the DNA, appears to be common to all these different proteins. In the case of the T4 and *E. coli* proteins, the exposed aromatics and their interaction with DNA were established by using NMR[22,23] and other techniques.[24] Biochemical data suggest that the gross structures of the proteins have little in common, as has already been pointed out in a review.[20] For example, the *E. coli* protein is a tetramer with 19 500-Da subunits, while the phage T4 protein is a dimer with 34 000-Da subunits. Also, the T4 protein binds zinc, while the *E. coli* and fd proteins have no metal requirement. However, a recent analysis of their amino acid sequences revealed a relationship.[25] In the analysis, it was assumed that the small fd protein might represent a single-stranded DNA binding motif that has the functional aromatic and basic residues in key positions. A search of the other sequences revealed regions that display a similar pattern of these residues and might, therefore, adopt the fd structure.

B. Ribonuclease (RNase)

Ribonuclease is a monomeric protein of molecular weight 13 800 Da (124 amino acids) that cleaves RNA at sites 3′ to pyrimidine residues.[26] The structure of the enzyme has been well characterized (for a review, see ref 27), and the active site, which involves histidines-12 and -119 and lysine-41, is well understood. RNase can also bind DNA, but the absence of the 2′OH prevents cleavage. The preference of the enzyme for single-stranded nucleic acids gives it primitive DNA unwinding properties.[28]

There are data indicating that RNase binds a much larger segment of nucleic acid than that which occupies the active site (dinucleotide). It was demonstrated that up to 11 nucleotides are protected by the binding and that at least 7 ion pairs are created in the process.[29] Brayer and McPherson[30] were able to cocrystallize RNase with the oligonucleotide $d(pA)_4$, and the structure of the complex shows how the protein is able to immobilize the single-stranded nucleic acid and direct it to the active site.[29,31]

The asymmetric unit of the crystal consists of 1 RNase and 4 oligonucleotides. The conformation of the protein is identical, within error, with that determined previously, although many of the surface side chains, especially those in contact with the oligonucleotides, are in different orientations. Of the 16 nucleotides in the complex, only 12 are bound directly to the protein, and these form a path that runs in a consistent 5′ to 3′ direction across the surface of the protein (see Figure 3). The active site is toward the 3′ end of the path, and the 5′ end is close to a patch of positive charges on the opposite side of the protein. It has been proposed that this patch, which consists of lysine and arginine residues, is an electropositive "ionic trap" that may serve as an initial binding region for the RNA molecule.[32] The path can be regarded as a line of lysine and arginine residues that leads the nucleic acid by a series of salt bridges from the ionic trap to the active site. The path is entirely nonspecific as regards the nucleic acid, and the only bases that contribute to the binding are those close to the active center. It is further speculated that the constellation of lysine and arginine residues

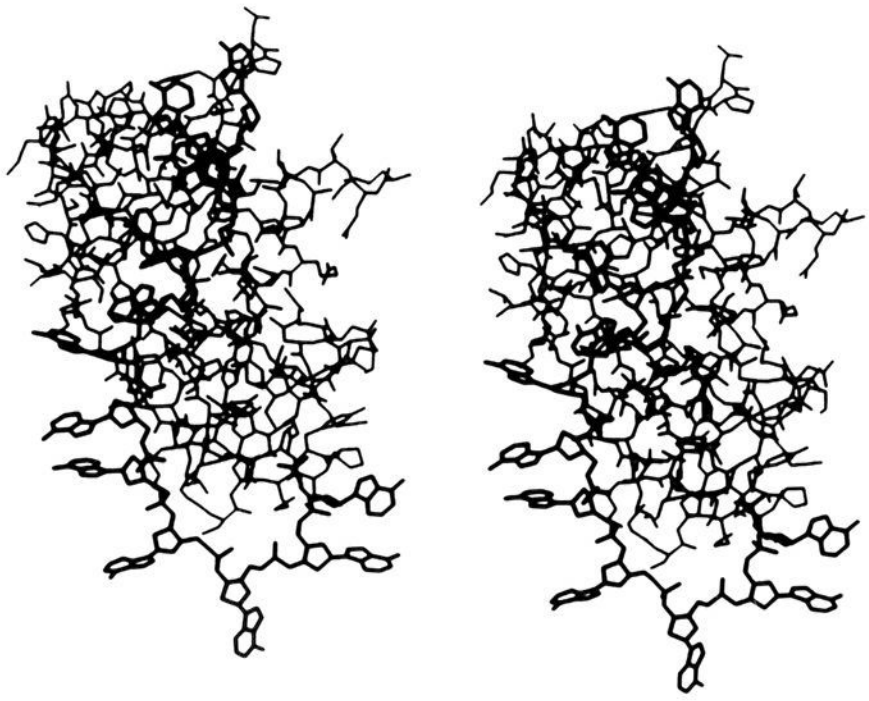


Figure 3. Course of the polynucleotide chain complexed to the RNase molecule. The path consists of the 12 nucleotides that are bound directly to the protein in a complex between RNase and 4 tetranucleotides. Reprinted with permission from ref 31. Copyright 1986 AAAS.

may complement the natural conformational preferences of the RNA and thereby lower the free energy of the complex.

C. The Eukaryotic Nucleosome

In eukaryotes, the large genomic DNA molecules are efficiently packaged into a structure called chromatin. Small basic proteins called histones form the basis of this structure. There are five types of histone protein called H1, H2A, H2B, H3, and H4. Two copies of each of the latter four proteins combine to form an octomeric core particle around which the DNA is supercoiled, and H1 binds on the outside (for a review, see ref 33). Core particles lacking H1 and linking DNA (DNA between core particles and not in the complex) have been purified and crystallized.[34] The structure is presently at 7 Å resolution.[35]

The complex has a disklike structure 57 Å thick and 110 Å in diameter. The DNA can be clearly seen on the outside, and it forms 1.8 turns of a left-handed superhelix of pitch 28 Å. The DNA is in the right-handed B form, and the superhelix is created by a series of kinks rather than by a smooth curve. The protein core contains a tetramer of $(H3{\cdot}H4)_2$ at the center and a (H2A·H2B) dimer on either side. DNA–protein interactions only occur on the inner surface of the superhelix and cannot be described in any detail at 7 Å resolution. However, the H3 dimer within the $(H3{\cdot}H4)_2$ tetramer appears to bind the DNA in a manner reminiscent of the small regulatory DNA binding proteins such as lambda repressor (see below). Two rods of density from each H3, which are probably α helixes, make contact with the minor grooves of successive turns of the DNA rather than with the major grooves.

A medium-resolution model of the histone octomer in the absence of DNA has been reported.[36] This shows major inconsistences with the nucleosome structure just described, and these have yet to be satisfactorily resolved.[37]

D. DNA Binding Protein II (HU)

The discovery of chromosome-like structures in prokaryotes[38] prompted the search for histone-like DNA binding proteins in these organisms. Several possible candidates were found (for a review, see ref 39), including a small basic protein of molecular weight 9500 Da (90 amino acids) referred to variously as HU, NS, and DNA binding protein II. It was shown that the purified protein can compact the contour length of DNA and induce the formation of beadlike structures in a way similar to the histone proteins.[40] Other results suggest that the DNA is negatively supercoiled in these complexes.[41] There are many copies of the protein in the cell (between 2×10^4 and 10^5), and it has been found in all the major bacterial groups. It can bind nonspecifically to single- and double-stranded DNA, and to RNA, and the abundant amino acid sequence information show it to be highly conserved. The protein from *Bacillus stearothermophilus* has been crystallized[42,43] and the structure determined.[44]

The protein normally exists as a dimer, and each monomer is in two distinct halves (Figure 4). The N-terminal half consists of two α helixes connected by a broad turn, which creates a vee-shaped motif similar to that found in the DNA-binding regulatory proteins (see below). The C-terminal half is a three-stranded antiparallel sheet followed by a short helix. In the dimer, the two helical halves wrap around each other to form a wedge-shaped base, and this is covered by the two sheets. The second and third strands of each sheet

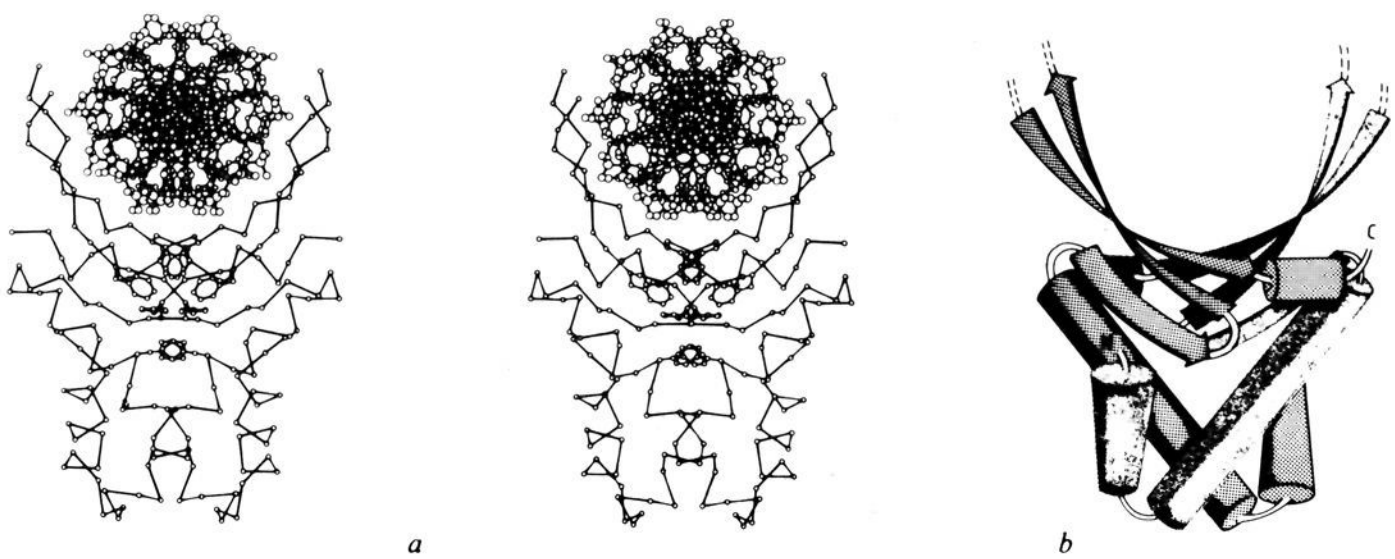


Figure 4. Structure of the DNA binding protein II: (a) stereo drawing of the α carbon backbone with the putative interaction with DNA, (b) Schematic representation of the secondary structure with each monomer shaded differently. Reprinted with permission from ref 44. Copyright 1984 Macmillan Journals Limited.

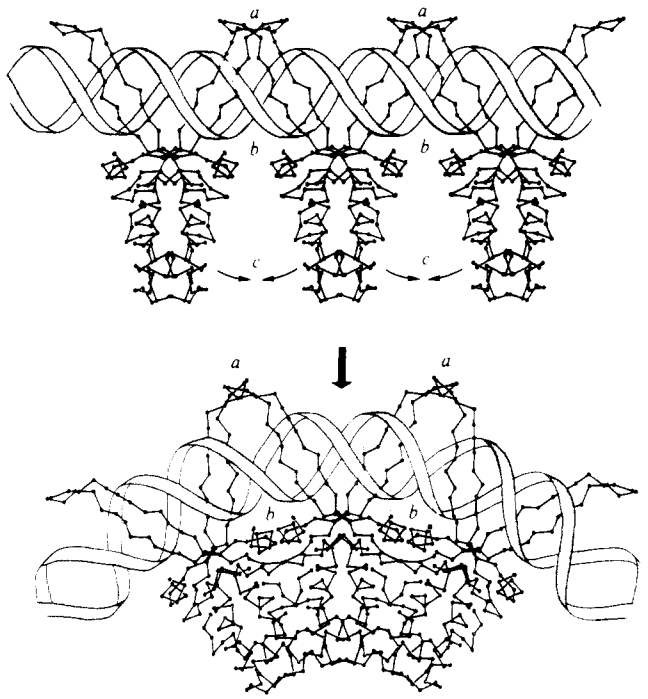


Figure 5. Proposal for how DNA binding protein II induces a supercoiling of the DNA. Likely conformations of the distal arms (unseen in the electron density map) are included. The important interactions are labeled (a) contact between the ends of the arms of adjacent molecules, (b) contact between the C-terminal helixes, (c) interactions between the wedge-shaped molecules. Reprinted with permission from ref 44. Copyright 1984 Macmillan Journals Limited.

are connected by a 26-residue segment that, for the most part, is invisible in the crystal structure. This is referred to as the arm region.

The visible parts of the arms and the two surfaces of the β sheets create a concave region of the protein that exactly complements the right-handed B-DNA helix (Figure 5). The sequence of the *B. stearothermophilus* protein reveals seven highly conserved basic residues in this region. These are located in the first half of the arm, the third strand of the sheet, and the short C-terminal helix. The visible, proximal part of the arm has a β ribbon-type structure, and if this is continued, a plausible DNA binding model can be constructed in which the basic residues form ionic interactions with the DNA phosphates. NMR studies showed that arginines are involved in the DNA binding.[42,45] The natural twist of the ribbon together with two bends in the arm at conserved proline residues would ensure that the arm follows the DNA helix. The model makes no predictions concerning the position of the arm relative to the major and minor grooves or indeed whether the arm penetrates the grooves at all.

A particularly attractive feature of the model is that it suggests how the protein might supercoil the DNA into a nucleosome type of structure. In the model, one molecule binds one turn of the DNA, and when it is viewed normal to the DNA helix axis, the wedge shape of the protein is particularly pronounced. One can therefore imagine that several molecules bound sequentially along a stretch of DNA could self-associate into a circular object with the DNA bound on the outside (Figure 5). It can be estimated that the resulting structure would be similar to one half of the eukaryotic nucleosome.

Two other proteins, TF1[46] and IHF,[47] were found to have amino acid sequences similar to DNA binding protein II and almost certainly have the same three-dimensional structure. TF1 (SP01 transcription factor 1) is coded for by the bacteriophage SP01 and is thought to help segregate viral DNA from that of the infected cell. It has a partial specificity in that it preferentially binds to SP01 DNA, which contains 5′-hydroxymethyluracil instead of thymine. It has a phenylalanine residue in place of a totally conserved arginine at the end of the arm (residue 61), and this may have a role in the partial specificity.[48] IHF (integration host factor) consists of two different subunits (a and b) and is coded for by *E. coli*. Although its precise function is not known, it binds specifically to DNA and has been implicated in the regulation of several bacterial genes. The molecular basis of the DNA recognition is not known.

E. Deoxyribonuclease I (DNase I)

This protein is an endonuclease that degrades double-stranded DNA to yield 5′-oligonucleotides. Although not a specific endonuclease, it was shown to cut preferentially certain types of DNA sequences.[49] The structure of bovine pancreatic DNase I has been solved to high resolution, and the binding of calcium deoxythymidine-3′,5′-diphosphate (Ca-pTp) at the active site has been investigated.[50]

The molecule has a molecular weight of 30 000 Da (257 amino acids) and contains a carbohydrate chain attached through an asparagine side chain. The core of the protein consists of a sandwich of two six-stranded β pleated sheets, and this is surrounded by α helixes and several extensive loop regions. Both sheets are antiparallel apart from the two central strands, and are related by a quasi-twofold symmetry axis. The loop regions are stabilized by intramolecular hydrogen bonds, salt bridges, two calcium binding sites and two disulfide bridges.

Investigations into the reaction mechanism (see ref 51) clearly showed the role of histidine-131 as a general base. Also, the scissile phosphorus experiences an inversion of configuration during the cleavage, which reflects a single displacement step. The reaction also has an absolute dependence on divalent metal ions (normally calcium). The crystal structure of the enzyme and the location of the Ca-pTp substrate suggest a mechanism of action.[51,52] The Ca-pTp binds in a shallow groove close to the N and C termini of the strands at the edge of the β sheet sandwich. The catalytic histidine-131 is situated between a glutamic acid residue and a water molecule, and these appear to form a charge relay system of the type found in the serine proteases.[11] The calcium ion is thought to have two crucial roles. It holds the scissile phosphate in the correct orientation at the active site and facilitates the nucleophilic attack by the water molecule.

This geometry at the active site together with the distribution of nearby charged residues and the shape of the surrounding protein surface immediately suggested how DNase I interacts with double-stranded DNA.[51,52] Central to the model (see Figure 6) is a loop that binds into the minor groove and salt bridges to the phosphates of both DNA strands. Two arginines are close to the two phosphates on the 5′ side of the scissile

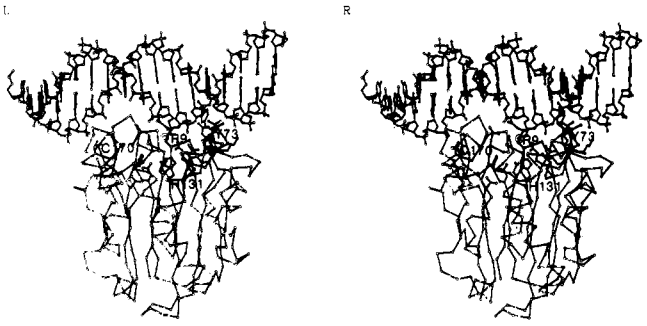


Figure 6. Stereo drawing showing the model for DNase I–DNA interaction. The DNA and side chains of DNase I in contact regions are drawn in continuous lines. The double circle marks the catalytic calcium. The exposed loop Arg-70-Asn-71-Ser-72-Tyr-73-Lys-74 interacts with the minor groove of the DNA. Reprinted with permission from ref 52. Copyright 1986 Academic Press Inc. Ltd.

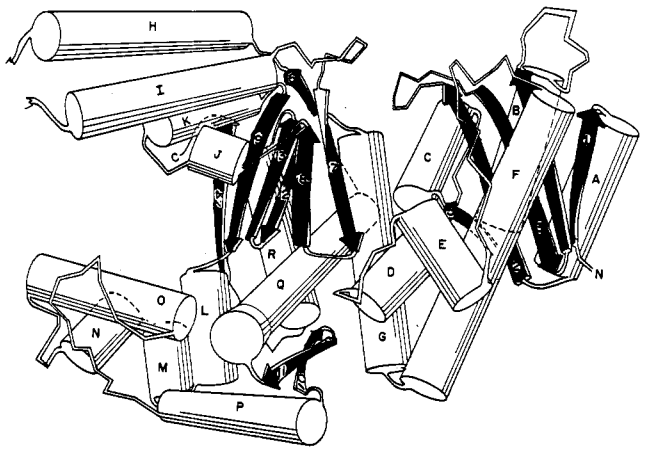


Figure 7. Schematic representation of the Klenow fragment. The cleft that runs through the center of the C-terminal domain is the putative DNA binding cleft.

phosphate, and another arginine contacts the phosphate on the 3′ side. Across the minor groove, phosphates on the other strand interact with a lysine and arginine. Nearby glutamic and aspartic acid side chains may also form ionic interactions with the DNA backbone through divalent cations. The model also predicts that van der Waals-type interactions further away from the active site may be important in aligning the DNA. Finally, the side chain of a tyrosine penetrates the minor groove and can potentially H bond to either the O2 of pyrimidines or the N3 of purines.

The proposed model differs from other protein–DNA interactions in that it involves a tight association between the protein and the minor groove of the DNA. There are, however, independent data to support this. The antibiotic netropsin is known to bind in the minor groove in A + T rich regions of DNA,[53] and it protects such regions from DNase I degradation. It is also known that DNase I cleaves self-complementary oligonucleotides in the middle and at the 3′ end more readily than at the 5′ end.[54] The model shows that in the latter case the DNA has to be aligned in a way that results in minimal protein–minor groove interaction. Chemical modification experiments indicated a role for the tyrosine, which, in the model, is located in the minor groove (see ref 51).

The model provides two possible explanations for the partial specificity of the enzyme. First, the proposed interaction would be hindered by a smaller than average minor groove, and this feature of duplex DNA is known to vary considerably with the DNA sequence.[8,9] Second, the tyrosine in the minor groove would experience considerable steric hindrance from the amino group in the 2-position of a guanine in G + C base pairs.

A particularly interesting feature of the model is a complementarity between the left-handed twist of the central sheet sandwich and the right-handed twist of the DNA double helix. More specifically, the residues involved in the binding and cleavage of the scissile strand of the DNA are situated at the N or C termini of the strands and in the connecting loops, and therefore follow the path of the DNA.

Recently, cocrystals of DNase I and self-complementary oligonucleotides were obtained, and the structural analysis of one of these confirmed the basic features of the model.[52]

F. The Klenow Fragment

The structure of the large fragment of DNA polymerase I from *E. coli* was determined at 3.2 Å resolution.[55,56] The intact DNA polymerase I is commonly referred to as Pol I, while the large fragment, first produced by proteolysis,[57,58] is known as the Klenow fragment. Pol I has three enzymatic activities: a DNA polymerase, a 3′ to 5′ exonuclease (editing), and a 5′ to 3′ exonuclease activity. The latter is not present in the Klenow fragment. Pol I is the best characterized of all DNA polymerase molecules.[58] It is known that Pol I will not bind to closed circles of double-stranded DNA and will only bind weakly to single-stranded DNA. However, it will bind tightly to a nick in double-stranded DNA.[58] It is also known that Pol I will bind dNMP (the product of the 3′ to 5′ exonuclease activity) and dNTP (a substrate for the polymerase activity).[58]

In *E. coli*, Pol I functions as part of the DNA repair system and is also used for the processing of Okazaki fragments.[58] Although Pol I is not the main replication enzyme in *E. coli*, the activities it possesses are very similar to those found in the main replication molecules.[58] The wealth of biochemical, genetic, and structural data available for Pol I make it the ideal system with which to study the detailed mechanisms of DNA replication.

As can be seen in Figure 7, the Klenow fragment folds into two domains, a small N-terminal domain of about 200 residues and a large C-terminal domain of about 400 residues.[55] The N-terminal domain fits into a structural class known as the Rossmann fold[11] and consists of a central twisted plane of β sheet with α helixes on either side. The connectivity of the secondary structure elements in the small domain has not been observed in other proteins.

The C-terminal domain of the Klenow fragment, shown in Figure 8, contains a cleft that is large enough to accommodate double-stranded B-DNA. Independent data from sequence comparisons,[59] electrostatic calculations,[60] chemical cross-linking,[59] and DNase I footprinting[61] are all consistent with the idea that the cleft is indeed used by Pol I to bind DNA. The DNA polymerase activity is processive, that is, it binds to a suitable primer terminus and fixes a number of bases before it dissociates.[58] The processive nature of polymerase molecules is clearly important for their func-

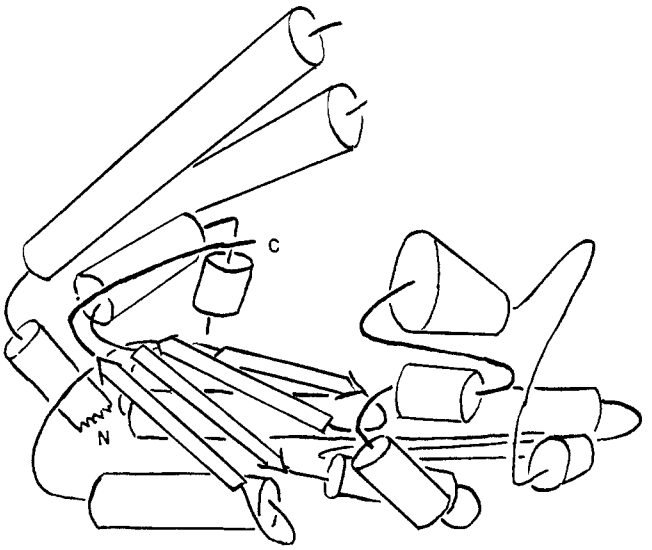


Figure 8. Schematic drawing of the C-terminal domain of the Klenow fragment. The viewer is looking down the DNA binding cleft.

tion and is probably the functional property that most clearly distinguishes Pol I from the main bacterial replication enzyme Pol III. The structure of the Klenow fragment and the putative site of DNA binding give some insight into this aspect of polymerase activity. A disordered peptide exists above the DNA binding cleft between helixes H and I. Although invisible, this peptide is potentially capable of blocking the DNA's access to the cleft and would be required to swing away from the protein to allow the DNA to bind. Once the DNA is in place, this peptide could bind to the DNA and thus lock it in the binding cleft. The disorder of the peptide in the crystal is probably a result of its natural flexibility, which clearly supports such a mechanism. The helixes J and K protude into the DNA binding cleft and place restrictions upon how a model of DNA can be docked in the cleft. Optimal fit of the DNA in the protein cleft is achieved by placing the helical structure in the major groove of DNA so that the protein cannot translate freely along the DNA but is forced to follow the helical symmetry of the DNA. This simple structural feature will allow the polymerase to follow the primer terminus. Thus, the available structural data give considerable insight into how Pol I interacts with DNA.

The polymerase active site was located by photoaffinity labeling of 8-azido-dATP at the dNTP binding site of the protein.[61] This is situated at the base of and on the N-terminal side of the DNA binding cleft. dNMP binds to the small domain of the crystalline Klenow fragment in a region that is thought to be the active site for the 3′ to 5′ exonuclease (editing) activity. A similarity between the exonuclease active site of the Klenow fragment and that of staphylococcal nuclease was noted.[55] The relative locations of these sites, which are nearly 30 Å apart (in the absence of DNA), is an important feature of the molecule. Intuitively it would be expected that the editing activity would be in close proximity to the polymerase activity. This leads to the speculation that the two active sites are brought together in the active molecule by a structural change triggered by the binding of DNA. The two domains are connected by a single peptide strand than could function as a hinge to allow the DNA-dependent repositioning of the two active sites. This idea is reasonable on the basis of the present model, although it needs confirmation by analysis of a structure of the protein–DNA complex.

Although the question of the relative locations of the two active sites cannot be addressed with certainty at the present time, there is information regarding the closely related question of whether the small domain is necessary for polymerase activity. In other words, is the small domain necessary for the formation of an active protein–DNA complex? A clone that produces the large domain of the Klenow fragment was obtained and found to have DNA polymerase activity,[62] thus demonstrating that the small N-terminal domain is not required for DNA polymerase activity.

IV. Specific Protein–Nucleic Acid Interactions

A. DNA Binding Regulatory Proteins

Several years ago, the structures of three proteins that regulate gene expression at the level of transcription were determined. The structures of the catabolite gene activator protein (or CAP) from *E. coli*[75] and the cro gene product (or cro) from bacteriophage lambda[72] were the first to appear, and these were followed by the cI gene product (or lambda repressor) also from lambda.[64] Although each structure is quite different, they share a common two-helix motif that, on the basis of model building, was proposed to mediate the binding to DNA. Based on amino acid sequence analysis, this motif was predicted to occur in other similar proteins, and one prediction was confirmed by the subsequent structural analysis of the trp repressor from *E. coli*.[81] Recently, the common DNA binding model was confirmed by the direct visualization of a repressor–DNA complex.[68]

1. Lambda Repressor

Bacteriophage lambda encodes two repressor proteins that compete for and bind operator regions on the phage DNA. Lambda repressor is necessary for the maintenance of the lysogenic state, and the cro protein (see below) is necessary during the lytic cycle. The system represents a simple biological switch mechanism that controls the nature of the lambda infection.[63]

The molecule contains 236 amino acids and has two domains. The N-terminal domain binds the DNA operator, and the C-terminal domain promotes the formation of dimers that bind more tightly to the operator sites. The N-terminal domain was isolated after proteolytic digestion, crystallized, and solved to high resolution.[64] It consists of an extended arm and five α helixes (numbered 1–5). In the crystal, the fragment forms a dimer through hydrophobic contacts.

The mode of interaction with DNA was deduced by model building, and this was considerably helped by biochemical and genetic data (see ref 64). These indicated that the protein binds to one face of the DNA in the major groove. Also, it was clear that the protein binds to operator DNA as a dimer and the complex has an approximate 2-fold symmetry. Finally, it was known that the three residues at the N terminus contact bases at the back of the operator and must therefore wrap around the DNA helix.[65] The most plausible model of DNA binding is based on the assumption that the di-

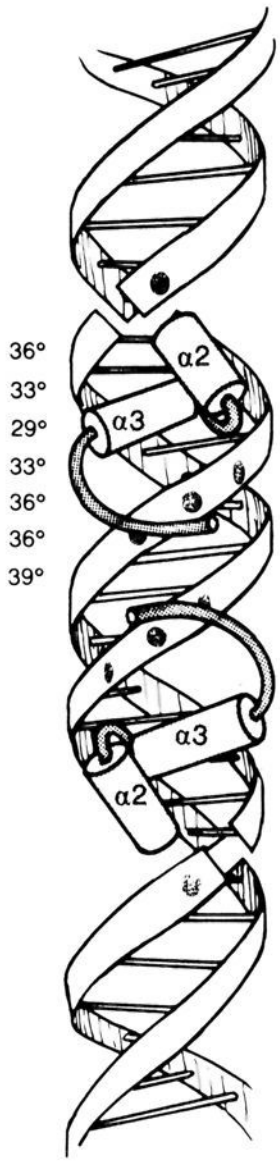


Figure 9. Diagram summarizing the DNA conformation and backbone contacts in the 434 repressor–operator complex. The numbers on the left indicate the twist in degrees between the base pairs. Reprinted with permission from ref 68. Copyright 1987 Macmillan Journals Limited.

meric form found in the crystal is very similar to that which binds to the operator. The central feature is a pair of symmetry related α helixes (helix 3), which are oriented such that their N termini point directly into successive turns of the major groove. The side chains of glutamine and serine residues at the ends of the helixes are well positioned to form hydrogen bonds to the bases. This model superimposes the protein and operator 2-fold rotation axes.

2. 434 Repressor–Operator Complex

Coliphage 434 codes for a repressor and a cro protein that operate a molecular switch very similar to that of bacteriophage lambda (see above) that decides between lysogeny and lytic growth. The DNA binding domain of the repressor (the first 69 residues of the complete protein referred to as R1-69) was crystallized with a 14-base-pair, self-complementary, synthetic operator DNA.[66] The structure was initially reported at low (7 Å) resolution,[67] and although little molecular detail was visible, it provided the first direct evidence that the common model of DNA binding proposed for lambda repressor, cro, CAP, and trp repressor was essentially correct. The complex contains a dimer of R1-69 bound to the operator, and two symmetry-related helixes are clearly visible in successive turns of the major groove. The structure has now been determined to 3.2 Å resolution.[68]

The protein monomer is a bundle of four helixes (H1 to H4), and a nonhelical C-terminal extension. The

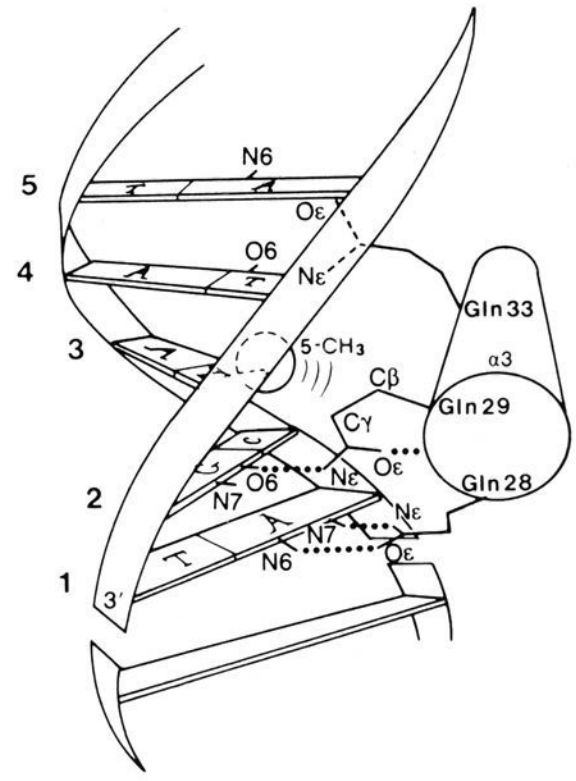


Figure 10. Schematic illustration of the interaction between α helix 3 of 434 repressor and DNA. The side chains of protein residues making contact with the DNA are shown. Reprinted with permission from ref 68. Copyright 1987 Macmillan Journals Limited.

helixes are arranged in a conformation very similar to the first four helixes of lambda repressor, and H2 and H3 form a helix-turn-helix motif as predicted from the amino acid sequence.[69] The DNA is in the B-type, double-helical conformation and is packed in the crystal to form a nearly perfect continuous helix. The local twist varies (Figure 9) such that it is slightly overwound at the center (~39°/base pair) and slightly underwound at the ends (~29°/base pair). Also, the helix axis is somewhat bent, most noticeably between base pairs 4 and 5 and 10 and 11.

In the complex, the N termini of H2, H3, and H4 point toward the DNA, and H3 lies in the major groove. A loop between H3 and H4 follows the course of the DNA backbone. The specificity of the protein–DNA interaction is the result of three contributing factors:[68] a. Rigidly defined interactions between the protein and the DNA backbone that determine their relative orientation. b. Specific interactions between H3 side chains and the DNA bases. c. Monomer–monomer interactions dependent upon the local overwinding at the center of the DNA helix.

One monomer contacts four phosphate groups (Figure 9), one of which is actually a part of the adjacent complex in the crystal but is close to the position that would be occupied in a continuous helix. These interactions involve both the side chains of basic residues and several main chain NH groups and are consistent with phosphate ethylation experiments (see ref 68). Two particularly interesting observations were made.[68] First, the distortions in the DNA structure result in a narrowing of the minor groove so that an arginine side chain inside the minor groove can form links with phosphates on either side. Second, the use of main chain amide groups in contact with the DNA backbone is thought to provide a more exact binding than could be achieved with the potentially more mobile side chains, and this is thought to be essential for the correct orientation of the side chains that interact with the bases.

Three glutamine residues project out of helix H3 and interact specifically with the bases in the major groove (Figure 10). The interactions are principally hydrogen bonds, but there is one clear van der Waals contact between the Cβ and Cγ atoms of one glutamine side chain and the 5-methyl group of a thymine. The importance of the interaction of 2 of the residues is reflected in the highly conserved nature of base pairs to which they bind in all 12 operator half-sites and the observation that there are no viable mutations of these amino acids.[68]

Since each monomer is firmly attached to the DNA, the formation of the dimer depends critically on the local twist of the helix. In this case, the DNA has to be overwound to ensure the correct monomer–monomer contacts. This seems to be confirmed by the discrimination displayed by the repressor against GC and CG base pairs at positions 6–9.[70] It is known that these bases are not easily accommodated in a narrow minor groove[71] and would therefore tend to prevent overwinding.

3. Lambda cro Protein

This protein has a molecular weight of 7351 Da and contains 66 amino acids. In the crystal,[72] there are 4 monomers in the asymmetric unit and the extent of their interaction suggests that cro may be a tetramer in solution. Each monomer contains a three-stranded antiparallel β pleated sheet and three helixes (numbered 1–3). As with the lambda repressor, the mode of DNA binding was investigated by model building using a considerable body of independent data (see ref 72). These show that lambda repressor and cro bind to the operator in a very similar way. In this case, helix 3 protrudes from the protein surface and, with a 2-fold partner, forms a surface feature that can interact with successive turns of the major groove.

This model was more thoroughly investigated by a combination of computer graphics and energy minimization.[73] The final model showed that the main interactions involve helixes 2 and 3, the third β strand, and the carboxy terminus. Helix 3 completely penetrates the major groove, and its polar and charged side chains make specific hydrogen bonds to the bases. Some of these are bidentate in nature, i.e., they involve multiple interactions between a side chain and more than one base pair. Helixes 2 and 3 appear also to bind nonspecifically and may make up to five hydrogen bonds to the phosphate groups. A further feature of the model is that two symmetry-related tripeptides form a central antiparallel β ribbon that is in a position to interact with the minor groove of the operator in a manner similar to that predicted by model building.[14] Finally, there also appear to be a number of hydrophobic interactions between protein side chains and the C7 methyl groups of thymines.

4. Catabolite Gene Activator Protein (CAP)

CAP forms a complex with the allosteric effector cyclic AMP. The complex then binds to specific DNA sites near the promotors of several operons and thereby alters their rate of transcription by RNA polymerase.[74] In *E. coli*, CAP is a positive regulator for the lactose and arabinose operons and can be either a positive or negative regulator for the galactose operon. The structure of the *E. coli* CAP–cAMP complex has been determined.[75,76] The active species is known to be a dimer, and each subunit has a molecular weight of 22 500 Da and contains 201 amino acids. The results of proteolytic digestion suggest that the molecule has two domains, and other studies indicate that a conformational change occurs when cAMP is bound (see ref 75).

The molecule, as predicted, has two domains: a larger N-terminal domain of 135 amino acids and a smaller C-terminal domain of 65 residues. The larger domain contains two helixes (A and B) and an eight-strand antiparallel β pleated sheet folded into a β barrel. The other domain is simply three helixes (D, E, and F). The domains are connected by a long (40 Å) helix C. The subunit interface in the dimer consists of the entire length of helixes C, and the β barrel of one and the central position of helix C from the other. There are no obvious contacts between the small domains. In the crystal, the dimer is the asymmetric unit, and the smaller domains are in different orientations with respect to the larger domains. This is the result of different conformations in a hinge region of four to five amino acids that connects the C helix to the small domain. cAMP is located in the interior of the β barrel.

Once again, model building and other data were used to investigate the interaction of CAP with DNA. The F helixes are separated by approximately 34 Å and protrude out of the protein surface. Their angle of tilt originally suggested that they bind in the major groove of left-handed B-DNA. However, the electric field of the protein was later shown to be complementary to right-handed B-DNA.[77] The present model for the interaction of the protein and DNA is similar to those for the lambda and trp systems (see below) and is consistent with genetic data[78] and other experiments.[79]

5. trp Repressor

In *E. coli* and other enteric bacteria, the biosynthesis of L-tryptophan is controlled by a simple negative feedback loop involving a small protein called the trp repressor.[80] When the concentration of L-tryptophan increases, this dimeric protein binds two molecules of L-tryptophan, which converts it from the aporepressor to the active repressor form. This then binds specifically to the operator sites of trp EDCBA (the L-tryptophan biosynthetic operon) and aroH (an aromatic amino acid biosynthetic operon) and shuts down their transcription. The structure of the active trp repressor has been determined.[81]

The molecule is almost entirely α helical with 72% of the residues contributing to six helixes (A–F), which are connected by short turns. In the dimer, these helixes are highly interwoven, and only the short helix D appears not to contribute to the extensive interface.

On the basis of sequence comparisons with other repressors, a region of the protein appeared likely to form a two-helix motif, which is central to the DNA binding models of lambda repressor, cro, and CAP (see below). This was confirmed by the structure and resulted in a model in which the N termini of symmetry-related helixes E point into the major groove. The seven N-terminal residues appear to be flexible and may interact with the side or back of the DNA.

Before the structure determination, the prediction of the two-helix motif prompted Kelley and Yanofsky[82]

to search for mutants in the trp R gene (which codes for the repressor) to confirm the idea. They found 9 different nonsense mutations (which produce truncated molecules) and 19 missense mutations (which produce single amino acid changes). These mutants differ in their ability to interfere with the wild-type protein when introduced into the cell. These differences are nicely reconciled by the structure and the DNA binding model. Of the nonsense mutations, only the one truncated at position 68 had a strong negative effect. The first 68 residues are sufficient to dimerize with the wild type, but would lack the DNA binding region. The negative complementing missense mutations occur either in the DNA binding region, or at positions surrounding the tryptophan binding pocket. The latter presumably interfere with the switching mechanism (see later). Finally, four superrepressors were found that exhibited repressor properties at very low tryptophan concentrations. Three of these involve a glycine to lysine mutation in regions of the model that face the DNA. The fourth involves an alanine to valine mutation at position 77. This is a region critical to the stability of the two-helix motif.

6. *The Helix-Turn-Helix Motif*

Central to all the models for the interaction between these small repressor proteins and their operators is a pair of symmetry-related α helixes that penetrate successive turns of the major groove. These provide the hydrogen-bond acceptors and donors for specific interaction with the bases and are referred to as recognition helixes. In cro, the helix is thought to lie in the groove, whereas in lambda repressor, CAP, and trp repressor, the N termini appear to point into the groove. The model of the lambda repressor–operator complex agrees well with the known structure of the closely related 434 operator complex. It has been pointed out[69] that the latter arrangement has two advantages. First, the net positive charge at the end of the helix[83] would be close to the negatively charged DNA, and second, the side chains naturally point toward the N terminus[11] and therefore toward the DNA. The structural similarity extends to a second helix, which immediately preceeds the recognition helix in the sequence. Together, they form a vee-shaped structure in which the inter-helix angle is constant. A search of all known protein structures revealed that this motif is almost unique to these proteins (for a review, see ref 69). It does occur in the nonspecific DNA Binding protein II (see above).

Before the model was definitively proved to be correct by the structure determination of the 434 repressor complex, there were many elegant genetic and biochemical experiments that supported it. Ptashne and co-workers succeeded in altering the specificity of 434 repressor to those of 434 cro protein[84] and P22 repressor[85] by site-specific mutagenesis. It was also possible to understand from the model how lambda repressor and lambda cro protein are able to distinguish between the six operator sites on the phage chromosome.[86,87] Other experiments supported the DNA binding models of lambda cro protein,[88,89] catabolite gene activator protein,[78] and trp repressor.[82]

The amino acid sequences in the two-helix motif region show some conservation. This is due to certain structural constraints imposed by the motif. It spans 20 residues, and position 9 normally contains a glycine that most easily adopts the conformation required at the turn. At positions 5 and 15 are small hydrophobic residues that are in van der Waals contact and appear to maintain the correct inter-helix angle. Position 18, which is also between the helixes, normally has a bulkier hydrophobic side chain. Finally, the general pattern of amino acids is dictated by a combination of the α-helical parameters and the need to expose one face of the recognition helix to the outside and to bury the opposite face in the protein core. This pattern of 20 residues has been found in many small repressor-type molecules and is thought to reflect the presence of the two-helix motif.[69]

One such prediction was confirmed by the determination of the trp repressor structure (see above). This structure also suggests that the mechanism by which L-tryptophan activates the aporepressor involves the motif. Position 18 in the motif has an atypical glycine, but the indole ring of the bound tryptophan occupies the space normally filled by a bulky hydrophobic side chain and appears to stabilize the motif. Also, the tryptophan carboxyl and ammonium groups appear to hydrogen bond to exposed side chains on the recognition helix. This may position them to form the correct hydrogen bonds to the DNA bases. The structure of the aporepressor is under investigation,[81] and this ought to clarify the switching mechanism.

B. DNA-*Eco*RI Endonuclease

The structure determination of the DNA-*Eco*RI endonuclease provided the first detailed view of a specific DNA–protein interaction.[90] The enzyme has a molecular weight of 31 065 Da, contains 276 amino acids, and is a dimer under physiological conditions. It recognizes the palindromic sequence G·C A·T A·T T·A T·A C·G and hydrolyzes the phosphodiester bond between the G and A in a well-characterized reaction that requires magnesium ions. An *Eco*RI methylase recognizes the same sequence, methylates the central adenines, and thereby protects nonforeign DNA from digestion (for a review of these systems see ref 91). The protein can bind DNA both specifically and nonspecifically, and the latter is thought to speed up the formation of the specific complex by a process of facilitated diffusion.[92] The crystals of the complex were obtained with a tridecameric oligonucleotide containing the recognition sequence, and magnesium ions were omitted to prevent enzyme turnover.[93]

The complex is a 2-fold symmetrical globular structure approximately 50 Å in diameter. The DNA is embedded in one side with the major groove almost totally buried and the minor groove completely exposed. An important feature, which was apparent at low resolution,[94] is a series of kinks in the DNA, one at the center and two at symmetry-related positions on either side. The authors describe these as neokinks to emphasize that they are induced by the bound protein. The crystal structure of a very similar stretch of DNA was determined[6] and the kinks are not present.

The central kink (type I neokink) occurs between the central A and T bases (Figure 11) and can be regarded as a relative unwinding of the two halves of the DNA by approximately 25°. An important consequence of

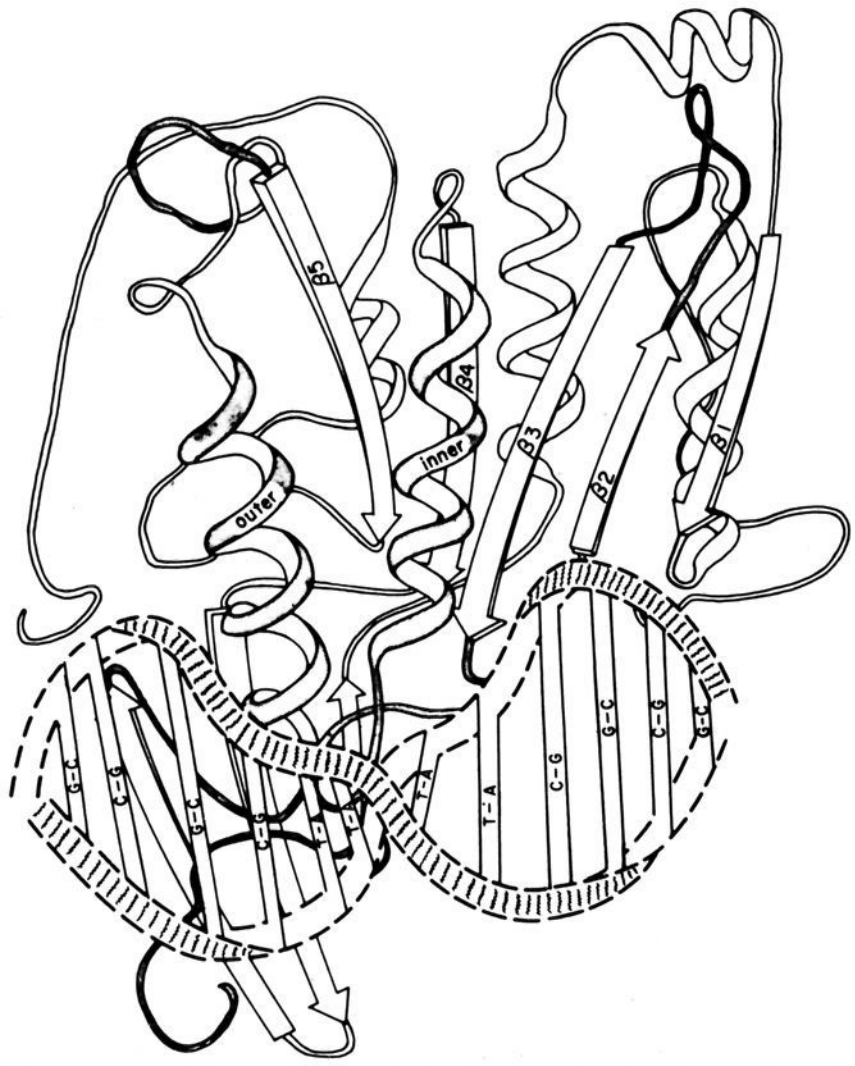


Figure 11. Schematic drawing of one subunit of the dimeric *Eco*R1 and both strands of the DNA complex. The helixes in the foreground of the diagram are the inner and outer recognition helixes. Reprinted with permission from ref 90. copyright 1986 AAAS.

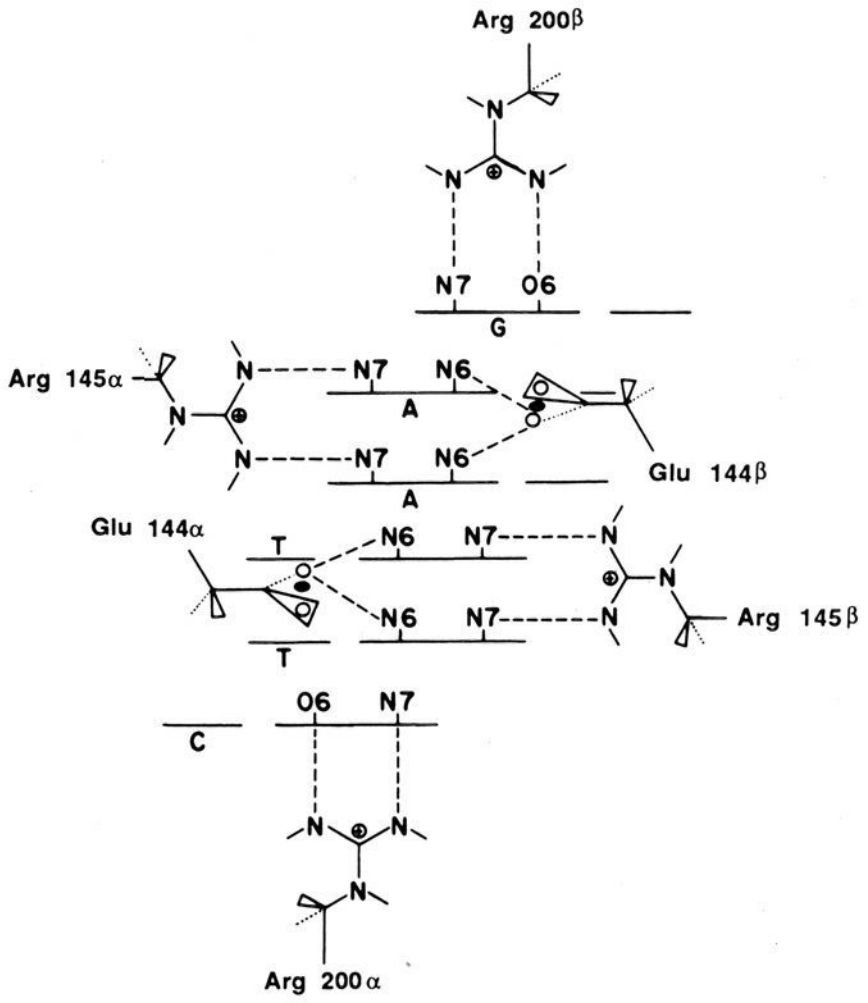


Figure 12. Schematic representation of the hydrogen-bond interactions that determine the specificity of the *Eco*R1 endonuclease. α and β refer to the two identical subunits of the enzyme. Reprinted with permission from ref 90. Copyright 1986 AAAS.

this kink as regards the complex formation is an expansion of the major groove by 3.5 Å as measured by the separation of the DNA backbones. This allows a greater penetration by the protein recognition units and may prove to be a general feature of these types of proteins. The outside kinks (type II neokinks) are centered at G4 and G10 (Figure 11) and appear to distort the scissile bond. These may be of importance to the enzyme mechanism.

The protein contains both β pleated sheets and α helixes and can be categorized as a typical α/β protein (Figure 11). Central to the structure is a five-stranded sheet in which strands 1, 3, 4, and 5 are parallel and strand 2 antiparallel. This sheet can be divided into two functional halves. Strands 1, 2, and 3 form an antiparallel sheet that appears to be involved in DNA scission. Strands 3, 4, and 5 form a parallel sheet that provides elements responsible for dimer formation and DNA binding. There are four long α helixes with their N termini pointing toward the DNA. Two of these helixes, referred to as the inner and outer helixes, point directly into the expanded major groove and their N termini carry the side chains that are responsible for DNA recognition.

In common with many other DNA binding proteins, *Eco*RI has an armlike structure that wraps around the DNA and holds it firmly with electrostatic interactions. The arm is composed of the N terminus and a three-stranded antiparallel sheet. It is located directly opposite the scissile bond in the region of the type II neokink (Figure 11). The arms may function to correctly orient the DNA in the active site and/or induce the kink formation. The fact that the DNA is completely encircled within the complex suggests that the free protein is in a more open configuration. Crystals have been grown in the absence of DNA, and their analysis should clarify this.[90]

The segments of the DNA backbone that are cleaved by the enzyme are buried in two symmetry-related clefts. The base of each cleft is formed by the edge of the antiparallel half of the main sheet and is close to phosphates 3, 4, and 5. The cleft is lined with many basic residues that interact with these phosphates. There is no obvious catalytic site, and it is suggested that the magnesium ion is required for its formation. Enzyme turnover does occur in the crystal after the addition of magnesium, and these are being studied in an effort to reveal the enzyme–product complex.[90]

The complex is formed by both nonspecific and specific protein–DNA interactions. The nonspecific interactions are extensive and involve DNA residues 2–9. The tight binding of the three phosphates in the catalytic cleft (3, 4, and 5) appears to distort the DNA helix and induce the type I neokink. Phosphates 8 and 9 interact with the arm region. These findings are consistent with phosphate ethylation interference experiments.[95] The specific recognition unit is a bundle of four parallel helixes, two each of the inner and outer helixes referred to above. The two inner helixes provide an arginine and a glutamic acid that form bridging hydrogen bonds to the central adenines of the recognition hexanucleotide. The arrangement is such that two adenines from one strand contact an arginine and glutamic acid from different subunits. The two outer helixes carry an arginine that forms bridging H bonds to the guanine (Figure 12). Note that the recognition only involves the purine bases. Note also that the arginines are involved in both inter- and intra-base

bridging interactions. This is a consequence of the different orientations of the inner and outer helixes. The former point directly at the helix axis, while the latter point toward the outside.

The high specificity of *Eco*RI can be relaxed under certain buffer conditions, and the protein is then able to cleave at other so-called *Eco*RI* sites.[96] It is thought that the formation of the catalytic site can only occur after formation of the recognition complex but that the abnormal conditions relax this requirement. The rate of cleavage of the *Eco*RI* sites can be directly correlated with the number of H bonds that can be formed.

The protein–DNA interface is characterized by a stable array of alternating positive and negative charges made up of DNA phosphates and protein side chains. The glutamic acid that is H bonded to the central adenines is at the center of this array, and these are the bases that are modified by the methylase to prevent complex formation. It is pointed out that the modification not only results in a loss of hydrogen bonds but also a disruption of the charge array by displacement of the glutamic acid side chain. It is also suggested that this electrostatic array may help the enzyme to discriminate against incorrect binding sites, since mismatches would disrupt the electrostatic interactions. There is evidence of a similar mechanism in the binding of lac repressor to its operator.[97]

C. Ribosomal Proteins

In every organism, genetic information is translated into protein by the ribosome and its associated cellular machinery. It is a nucleoprotein particle consisting of a large and a small subunit, and each subunit contains a large RNA molecule and a number of ribosomal proteins.[98] Approximately two-thirds of the mass of the ribosome is RNA, and protein–RNA interactions are likely to be crucial to the structure and function of the organelle. In addition, some of the ribosomal proteins exert feedback regulation on the translation of ribosomal protein mRNA.[99] They appear to recognize and bind regions of the mRNA that are closely similar to their binding sites on the rRNA. The RNA binding properties of the ribosomal proteins are reflected in their amino acid sequences, which contain many basic residues.[98] The high-resolution structures of two of these proteins have now been determined.[100,101]

L30[101] from the large subunit of the *B. stearothermophilus* ribosome is a small protein of molecular weight 6000 Da (60 amino acids). It comprises two helixes packed onto one side of a three-stranded antiparallel sheet. One end of the molecule has six isolated basic residues (four arginines and two lysines), which could interact with the ribosomal RNA. Also, four adjacent and conserved polar residues (three threonines and one serine) are well placed to form hydrogen bonds to the RNA. Isolated L30 does not bind to purified ribosomal RNA like some of the proteins, and it is likely to recognize a site on the partially assembled ribosome. For this reason, no attempts were made to produce a detailed model for the interaction.

The only other known ribosomal protein structure is that of the C-terminal fragment of L7/L12 from the large subunit of the *E. coli* ribosome.[100] This protein is known to be located on the ribosome at the distal end of a stalklike protuberance[98] and is relatively distant from the ribosomal RNA. This is reflected in the structure, which displays no obvious RNA binding regions.

D. Protein Transcription Factor IIIA (TFIIIA)

Protein transcription factor IIIA was shown to be essential for the synthesis of 5S RNA by RNA polymerase III in eukaryotes. It is known to bind to a stretch of DNA approximately 50 bp (bp = base pairs) in length within the coding region of the 5S RNA gene. The molecule from *Xenopus laevis* has been particularly well studied because it forms many copies of a complex with 5S RNA molecules in immature oocytes and is relatively easy to purify.[102] Although the three-dimensional structure of the protein has not been determined, the primary structure suggests a model of DNA binding that is supported by other evidence.

The protein has a molecular weight of 40 000 Da, and when it is subjected to proteolytic digestion, small 3000-Da fragments are generated, which suggests a repeating structure within the molecule.[103] An analysis of the primary structure revealed that about 70% of the protein consists of 9 tandemly repeated sequences, each of approximately 30 residues, and certain amino acids are highly conserved.[103] In particular, each segment contains a pair of histidines and cysteines. A similar analysis was performed on the DNA segment to which TFIIIA binds, and evidence was found that this also has a repeat structure every half-turn of the double helix. This was supported by quantitative DNase I digestion studies.[104] Finally, there was clear evidence that a stoichiometric amount of 7–11 zinc atoms is associated with each TFIIIA molecule.[103]

It is proposed[103] that each repeat unit forms a fingerlike structure that is stabilized by a zinc atom tetrahedrally coordinated to the conserved histidine and cysteine side chains at the base of the finger. The amino acids within the finger are generally of the correct type for binding to DNA. Since the conformation of the finger is unknown, no detailed description of its interaction with DNA has been possible. However, recent spectral data support the proposed coordination of the zinc atoms.[105] Also, it has been possible to map the sites on the 5S RNA gene that are protected by TFIIIA.[106] There are nine sites separated by about five base pairs, which agrees well with the analysis of the DNA repeat structure. The recent discovery of similar repeat regions in other regulatory proteins clearly shows that it is an important DNA binding motif.[107–109]

V. Conclusion

Oligomeric nucleic acids are highly complex polymers that serve a wide variety of functions and that interact with proteins for a multitude of reasons. It is now clear that nature has devised a range of protein architectures to deal with the nucleic acids. However, the predictions outlined in section II concerning the general modes of interaction between the two types of macromolecules have, for the most part, been confirmed. All the proteins contain strategically located positively charged residues that neutralize the negative charges on the nucleic acid and increase the binding affinity. Proteins that need to recognize a specific nucleic acid sequence do so through a combination of specific hydrogen bonds and hydrophobic interactions, and these almost always

occur in the major groove, which offers greater accessibility to the functional groups of the bases. The single-stranded binding proteins need to stabilize exposed bases, and this is achieved by the provision of aromatic side chains that participate in base stacking type interactions.

The predictions concerning the types of protein secondary structure that are involved in nucleic acid binding (see section II) have also been largely verified. Several structures and models feature α helixes that penetrate the DNA major groove, and other have antiparallel β ribbons close to the DNA backbone. Some of the proteins, however, have revealed additional general features. The *Eco*RI and DNase I structures both display a complementarity between the left-handed twist of a β pleated sheet and the right-handed twist of the DNA helix, and this may prove to be important in other nucleic acid binding proteins. Also, it is noticeable that the N termini of α helixes tend to point toward the nucleic acid rather than the C termini. This may contribute to the stability of the complex, since the N terminus of the helix has a net positive charge.[83] Finally, the structures of the two protein–duplex DNA complexes revealed considerable distortions in the DNA helix. Expansion of the major groove allows the protein recognition units to penetrate more deeply into the center of the DNA, and, in the case of *Eco*RI, the distortion of the sugar phosphate backbone may be an important part of the catalytic cleavage mechanism. It is also apparent from the 434 repressor complex that the DNA is bent to some degree around the protein. This has also been predicted for the lambda cro protein[73] and has been detected in other similar complexes.[110] The bending increases the interaction between the molecules and may serve an additional functional role.[110]

As regards the helix-turn-helix motif in the repressor molecules (section IV A), it was initially suggested[69] that this might be a universal DNA recognition element. The structure of the *Eco*RI complex clearly has no such motif, and it has been proposed that it represents an efficient method of DNA recognition for *small* proteins.[90] The first helix appears normally to straddle the major groove and acts as a stable platform for the recognition helix. The 434 repressor complex has hydrogen bonds between the phosphate groups and the rigid polypeptide backbone, and these firmly orient the protein on the DNA. A larger protein such as *Eco*RI has the potential for a more extensive contact with the DNA and could dispense with the first helix. The zinc binding fingers of TFIIIA may prove to be another small DNA binding motif.

The current interest in protein–nucleic acid interactions has been closely coupled with developments in molecular biology and genetic engineering. There have been numerous discoveries in molecular biology that have focused attention on protein–nucleic acid interactions, and these have provided the impetus for the rapid growth in structural investigations in this area. Of more practical importance has been the development of technology for the high expression of proteins. The quantities of proteins produced by these expression systems greatly facilitates the process of structure determination. Also, site-directed mutagenesis has and will continue to enable structure–function models to be tested. It is anticipated that the interaction of genetics and physical methods will increase in the future.

In the last few years the mode of interaction in some protein–nucleic acids systems has been revealed. Both physical and genetic approaches have been used to define these interactions so that a detailed understanding of the forces linking these molecules is starting to emerge. The near future should see an increased understanding of the systems reviewed here, and it is anticipated that the structures of other interesting nucleic acid binding proteins will emerge (for example, see ref 111–115).

Acknowledgments. We are grateful to Drs. J. J. Dunn, F. W. Studier, and J. Dubendorff for critically reading the manuscript and Dr. V. Ramakrishnan and Madeline Windsor for invaluable assistance. We also thank Drs. S. C. Harrison, M. Ptashne, and J. M. Rosenberg for supplying preprints of their research and Dr. A. McPherson for sending artwork. This work was in part supported by the Office of Health and Environmental Research of the United States Department of Energy.

VI. References

(1) Watson, J. D.; Hopkins, N. H.; Roberts, J. W.; Steitz, J. A.; Weiner, A. M. *Molecular Biology of the Gene* 4th ed.; Benjamin/Cummings: Menlo Park, CA, 1987 .
(2) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, Berlin, Heidelberg, Tokyo, 1984.
(3) Helene, C.; Lancelot, G. *Prog. Biophys. Mol. Biol.* **1982**, *39*, 1.
(4) von Hippel, P. *Biol. Regul. Dev.* **1979**, *1*, 279–348.
(5) Watson, J. D.; Crick, F. H. C. *Nature (London)* **1953**, *171*, 737.
(6) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179.
(7) McCall, M.; Brown, T.; Kennard, O. *J. Mol. Biol.* **1985**, *183*, 385.
(8) Calladine, C. R. *J. Mol. Biol.* **1982**, *161*, 343.
(9) Dickerson, R. E. *J. Mol. Biol.* **1983**, *166*, 419.
(10) Koo, H.-S.; Wu, H.-M.; Crothers, D. M. *Nature (London)* **1986**, *320*, 501.
(11) Creighton, T. E. *Proteins—Structures and Molecular Principles*; W. H. Freeman: New York, 1983.
(12) Zubay, G.; Doty, P. *J. Mol. Biol.* **1959**, *1*, 1.
(13) Carter, C. W., Jr.; Kraut, J. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 283.
(14) Church, G. M.; Sussman, J. L.; Kim, S.-H. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1458.
(15) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804.
(16) von Hippel, P. H.; Berg, O. G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1608.
(17) Coleman, J. E.; Oakley, J. L. *CRC Crit. Rev. Biochem.* **1980**, *7*, 247.
(18) Brayer, G. D.; McPherson, A. *J. Mol. Biol.* **1983**, *169*, 565.
(19) Brayer, G. D.; McPherson, A. *Biochemistry* **1984**, *23*, 340.
(20) Kowalczykowski, S.; Bear, D.; von Hippel, P. In *The Enzymes*; Boyer, P. D., Ed.; Academic: Orlando, 1981; Vol. 14a, p 373.
(21) Howard-Flanders, P.; West, S. C.; Stasiak, A. *Nature (London)* **1984**, *309*, 215.
(22) Prigodich, R. V.; Casas-Finet, J.; Williams, K. R.; Konigsberg, W.; Coleman, J. E. *Biochemistry* **1984**, *23*, 522.
(23) Prigodich, R. V.; Shamoo, Y.; Williams, K. R.; Chase, J. W.; Konigsberg, W. H.; Coleman, J. E. *Biochemistry* **1986**, *25*, 3666.
(24) Mustafa, I. K.; Casas, J. R.; Maki, A. H. *J. Biol. Chem.* **1987**, *262*, 1725.
(25) Prasad, B. V. V.; Chui, W. *J. Mol. Biol.* **1987**, *193*, 579.
(26) Blackburn, P.; Moore, S. In *The Enzymes*, 3rd ed.; Boyer, P. D., ed.; Academic: New York, 1982; Vol. 15b, p 317.
(27) Wlodawer, A. In *Biological Macromolecules and Assemblies, Nucleic Acids and Interactive Proteins*; Jurnak, F. A., McPherson, A., Eds.; John Wiley: New York, 1984, p 393.
(28) von Hippel, P. H.; Jensen, D. E.; Kelly, R. C.; McGhee, J. D. In *Nucleic Acid–Protein Recognition*; Vogel, H. J., Ed.; Academic: New York, 1977; p 65.

(29) McPherson, A.; Brayer, G. D.; Morrison, R. D. *J. Mol. Biol.* **1984**, *189*, 305.
(30) Brayer, G. D.; McPherson, A. *J. Biol. Chem.* **1981**, *257*, 3359.
(31) McPherson, A.; Brayer, G. D.; Cascio, D.; Williams, R. *Science (Washington, D.C.)* **1986**, *232*, 765.
(32) Matthew, J. B.; Richards, F. M. *Biochemistry* **1982**, *21*, 4989.
(33) Thomas, J. O. *J. Cell. Sci. Suppl.* **1984**, *1*, 1.
(34) Finch, J. T.; Brown, R. S.; Rhodes, D.; Richmond, T.; Rushton, B.; Lutter, L. C.; Klug, A. *J. Mol. Biol.* **1981**, *145*, 757.
(35) Richmond, T. J.; Finch, J. T.; Rushton, B.; Rhodes, D.; Klug, A. *Nature (London)* **1984**, *311*, 532.
(36) Burlingame, R. W.; Love, W. E.; Wang, B.-C.; Hamlin, R.; Xuong, N.-H.; Moudrianakis, E. N. *Science (Washington, D.C.)* **1985**, *228*, 546.
(37) Klug, A.; Finch, J. T.; Richmond, T. J.; Moudrianakis, E. N.; Love, W. E.; Wang, B.-C.; Xuong, N.-H.; Burlingame, R. W.; Uberbacher, E. C.; Bunick, G. J. *Science (Washington, D.C.)* **1985**, *229*, 1109.
(38) Griffith, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 563.
(39) Geider, K.; Hoffmann-Berling, H. *Ann. Rev. Biochem.* **1981**, *50*, 233.
(40) Rouvière-Yaniv, J.; Yaniv, M. *Cell* **1979**, *17*, 265.
(41) Broyles, S. S.; Pettijohn, D. E. *J. Mol. Biol.* **1986**, *187*, 47.
(42) Dijk, J.; White, S. W.; Wilson, K. S.; Appelt, K. *J. Biol. Chem.* **1983**, *258*, 4003.
(43) Tanaka, I.; White, S. W.; Appelt, K.; Wilson, K. S.; Dijk, J. *FEBS Lett.* **1984**, *165*, 39.
(44) Tanaka, I.; Appelt, K.; Dijk, J.; White, S. W.; Wilson, K. S. *Nature (London)* **1984**, *310*, 376.
(45) Lammi, M.; Paci, M.; Gualerzi, C. O. *FEBS Lett.* **1984**, *170*, 99.
(46) Greene, J. R.; Brennan, S. M.; Andrew, D. J.; Thompson, C. C.; Richards, S. H.; Heinrikson, R. L.; Geiduschek, E. P. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 7031.
(47) Flamm, E. L.; Weisberg, R. A. *J. Mol. Biol.* **1985**, *183*, 117.
(48) Greene, J. R.; Appelt, K.; Geiduschek, E. P. *Protein DNA Interaction and Gene Expression*; Texas University: Galveston, TX, in press.
(49) Moore, S. In *The Enzymes*; Boyer, P. D., Ed.; Academic: New York, 1981; Vol. 14, p 281.
(50) Suck, D.; Oefner, C.; Kabsch, W. *EMBO J.* **1984**, *3*, 2423.
(51) Suck, D.; Oefner, C. *Nature (London)* **1986**, *321*, 620.
(52) Oefner, C.; Suck, D. *J. Mol. Biol.* **1986**, *192*, 605.
(53) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553.
(54) Drew, H. R. *J. Mol. Biol.* **1984**, *176*, 535.
(55) Brick, P.; Ollis, D. L.; Steitz, T. *J. Mol. Biol.* **1983**, *166*, 453.
(56) Ollis, D. L.; Brick, P.; Hamlin, R.; Xuong, N. G.; Steitz, T. *Nature (London)* **1985**, *313*, 762.
(57) Klenow, H.; Henningsen, I. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *65*, 168.
(58) Brutlag, D.; Atkinson, M. R.; Setlow, P.; Kornberg, A. *Biochem. Biophys. Res. Comm.* **1969**, *37*, 982.
(59) Ollis, D. L.; Kline, C.; Steitz, T. *Nature (London)* **1985**, *313*, 818.
(60) Warwicker, J.; Ollis, D. L.; Richards, F. M.; Steitz, T. *J. Mol. Biol.* **1985**, *186*, 645.
(61) Joyce, C. M.; Ollis, D. L.; Rush, J.; Steitz, T.; Konigsberg, W. H.; Grindley, N. D. F. In *Protein Structure, Folding and Design*, UCLA Symposia in Molecular and Cellular Biology; Oxender, D., Ed.; Alan R. Liss: New York, 1986; p 197.
(62) Freemont, P. S.; Ollis, D. L.; Steitz, T.; Joyce, C. M. *Proteins Struct. Funct. Genet.* **1986**, *1*, 66.
(63) Ptashne, M. *Genetic Switch*; Cell: Cambridge, MA, 1986.
(64) Pabo, C. O.; Lewis, M. *Nature (London)* **1982**, *298*, 443.
(65) Pabo, C. O.; Krovatin, W.; Jeffrey, A.; Sauer, R. T. *Nature (London)* **1982**, *298*, 441.
(66) Anderson, J. E.; Ptashne, M.; Harrison, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1307.
(67) Anderson, J. E.; Ptashne, M.; Harrison, S. C. *Nature (London)* **1985**, *316*, 596.
(68) Anderson, J. E.; Ptashne, M.; Harrison, S. C. *Nature (London)* **1987**, *326*, 846.
(69) Pabo, C. O.; Sauer, R. T. *Ann. Rev. Biochem.* **1984**, *53*, 293.
(70) Koudelka, J.; Harrison, S. C.; Ptashne, M. *Nature (London)* **1987**, *326*, 886.
(71) Drew, H. R.; Travers, A. A. *Cell* **1984**, *37*, 491.
(72) Anderson, W. F.; Ohlendorf, D. H.; Takeda, Y.; Matthews, B. W. *Nature (London)* **1981**, *290*, 754.
(73) Ohlendorf, D. H.; Anderson, W. F.; Fisher, R. G.; Takeda, Y.; Matthews, B. W. *Nature (London)* **1982**, *298*, 718.
(74) Pastan, I.; Adhya, S. *Bacteriol. Rev.* **1976**, *40*, 527.
(75) McKay, D. B.; Steitz, T. A. *Nature (London)* **1981**, *290*, 744.
(76) McKay, D. B.; Weber, I. T.; Steitz, T. A. *J. Biol. Chem.* **1982**, *257*, 9518.
(77) Weber, I. T.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3973.
(78) Ebright, R. H.; Cossart, P.; Gicquel-Sanzey, B.; Beckwith, J. *Nature (London)* **1984**, *311*, 232.
(79) Kolb, A.; Buc, H. *Nucleic Acids Res.* **1982**, *10*, 473.
(80) Joachimiak, A. J.; Kelley, R. L.; Gunsalus, R. P.; Yanofsky, C.; Sigler, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 668.
(81) Schevitz, R. W.; Otwinowski, Z.; Joachimiak, A. J.; Lawson, C. L.; Sigler, P. B. *Nature (London)* **1985**, *317*, 782.
(82) Kelley, R. L.; Yanofsky, C. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 483.
(83) Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. *Nature (London)* **1978**, *273*, 443.
(84) Wharton, R. P.; Brown, E. L.; Ptashne, M. *Cell* **1984**, *38*, 361.
(85) Wharton, R. P.; Ptashne, M. *Nature (London)* **1985**, *316*, 601.
(86) Hochschild, A.; Ptashne, M. *Cell* **1986**, *44*, 925.
(87) Hochschild, A.; Douhan III, J.; Ptashne, M. *Cell* **1986**, *47*, 807.
(88) Eisenbeis, S. J.; Nasoff, M. S.; Noble, S. A.; Bracco, L. P.; Dodds, D. R.; Caruthers, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1084.
(89) Takeda, Y.; Kim, J. G.; Caday, C. G.; Steers, E.; Ohlendorf, D. H.; Anderson, W. F.; Matthews, B. W. *J. Biol. Chem.* **1986**, *261*, 8608.
(90) McClarin, J. A.; Frederick, C. A.; Wang, B.-C.; Greene, P.; Boyer, H. W.; Grable, J.; Rosenberg, J. M. *Science (Washington, D.C.)* **1986**, *234*, 1526.
(91) Modrich, P. *CRC Crit. Rev. Biochem.* **1982**, *13*, 287.
(92) Terry, B. J.; Jack, W. E.; Modrich, P. *J. Biol. Chem.* **1985**, *260*, 13130.
(93) Grable, J.; Frederick, C. A.; Samudzi, C.; Jen-Jacobson, L.; Lesser, D.; Greene, P. J.; Boyer, H. W.; Itakura, K.; Rosenberg, J. M. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1149.
(94) Frederick, C. A.; Grable, J.; Melia, M.; Samudzi, C.; Jen-Jacobson, L.; Wang, B.-C.; Greene, P. J.; Boyer, H. W.; Rosenberg, J. M. *Nature (London)* **1984**, *309*, 327.
(95) Lu, A.-L.; Jack, W. E.; Modrich, P. *J. Biol. Chem.* **1981**, *256*, 13200.
(96) Rosenberg, J. M.; Greene, P. J. *DNA* **1982**, *1*, 117.
(97) Mossing, M. C.; Record, Jr., M. T. *J. Mol. Biol.* **1985**, *186*, 295.
(98) Nierhaus, K. H. *Curr. Top. Microbiol. Immunol.* **1982**, *97*, 81.
(99) Nomura, M.; Gourse, R.; Baughman, G. *Ann. Rev. Biochem.* **1984**, *53*, 75.
(100) Leijonmarck, M.; Ericksson, S.; Liljas, A. *Nature (London)* **1980**, *286*, 824.
(101) Wilson, K. S.; Appelt, K.; Badger, J.; Tanaka, I.; White, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 7251.
(102) Engelke, D. R.; Ng, S. Y.; Shastry, B. S.; Roeder, R. G. *Cell* **1980**, *19*, 717.
(103) Miller, J.; McLachlan, A. D.; Klug, A. *EMBO J.* **1985**, *4*, 1609.
(104) Rhodes, D.; Klug, A. *Cell* **1986**, *46*, 123.
(105) Diakun, G. P.; Fairall, L.; Klug, A. *Nature (London)* **1986**, *324*, 698.
(106) Fairall, L.; Rhodes, D.; Klug, A. *J. Mol. Biol.* **1986**, *192*, 577.
(107) Vincent, A.; Colot, H. V.; Rosbash, M. *J. Mol. Biol.* **1985**, *186*, 149.
(108) Hartshorne, T. A.; Blumberg, H.; Young, E. T. *Nature (London)* **1986**, *320*, 283.
(109) Rosenberg, U. B.; Schroeder, C.; Preiss, A.; Kienlin, A.; Cote, S.; Riede, I.; Jaekle, H. *Nature (London)* **1986**, *319*, 336.
(110) Liu-Johnson, H.-N.; Gartenberg, M. R.; Crothers, D. M. *Cell* **1986**, *47*, 995.
(111) Weber, I. T.; Steitz, T. A. *J. Mol. Biol.* **1986**, *188*, 109.
(112) Brennan, R. G.; Takeda, Y.; Kim, J.; Anderson, W. F.; Matthews, B. W. *J. Mol. Biol.* **1986**, *188*, 115.
(113) Moras, D.; Lorber, B.; Romby, P.; Ebel, J. P.; Giege, R.; Lewit-Bentley, A.; Roth, M. *J. Biomol. Struct. Dyn.* **1983**, *1*, 209.
(114) White, S. W.; Appelt, K.; Dijk, J.; Wilson, K. S. *FEBS Lett.* **1983**, *163*, 73.
(115) Appelt, K.; White, S. W.; Wilson, K. S.; Tanaka, I. *FEBS Lett.* **1984**, *165*, 43.