*Annu. Rev. Biophys. Biomol. Struct. 1997. 26:289–325*




# EUKARYOTIC TRANSCRIPTION FACTOR-DNA COMPLEXES

*G. Patikoglou*
Laboratories of Molecular Biophysics, The Rockefeller University, New York, NY 10021

*S. K. Burley*
Laboratories of Molecular Biophysics and Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10021

KEY WORDS: eukaryotic transcription factor, protein-DNA complex, transcriptional activators, TATA box-binding protein, YY1, initiator element, DNA-binding domain

ABSTRACT

Eukaryotes have three distinct RNA polymerases that catalyze transcription of nuclear genes. RNA polymerase II is responsible for transcribing nuclear genes encoding the messenger RNAs and several small nuclear RNAs. Like RNA polymerases I and III, polymerase II cannot recognize its target promoter directly and initiate transcription without accessory factors. Instead, this large multisubunit enzyme relies on general transcription factors and transcriptional activators and coactivators to regulate transcription from class II promoters. X-ray crystallography and nuclear magnetic resonance spectroscopy have been used to study complexes of general transcription factors and transcriptional activators with their specific DNA targets. This work has provided important structural insights into transcription initiation by polymerase II and the more general problem of DNA sequence recognition.

CONTENTS





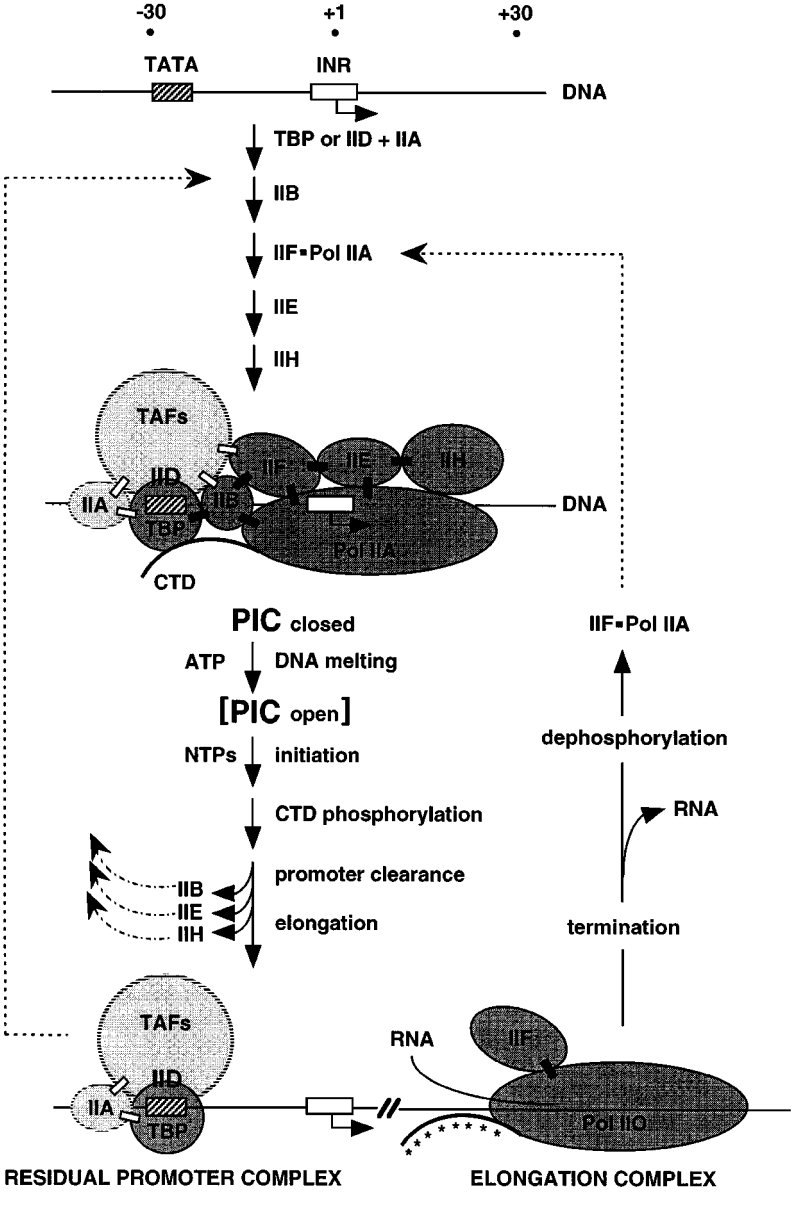
1056-8700/97/0610-0289$08.00

## INTRODUCTION

In eukaryotes, RNA polymerase II (pol II) is responsible for transcribing nuclear genes encoding the messenger RNAs and several small nuclear RNAs. Like RNA polymerases I and III, pol II cannot recognize its target promoter directly. Class II nuclear gene promoters contain combinations of DNA sequences, which include core or basal promoter elements, promoter proximal elements, and distal enhancer elements. Transcription initiation by pol II is precisely regulated by transcription factors (proteins) that interact with these three classes of DNA targets and also with each other (reviewed in 122). The best characterized core promoter elements, which can function independently or synergistically, are the TATA element, located 25 base pairs upstream of the transcription start site, and a pyrimidine-rich initiator (Inr) element, located at the start site. Individual core promoters may contain both, only one or the other, or neither of these elements. Promoter proximal elements occur anywhere between 50 and 200 base pairs upstream of the cap site (+1 in Figure 1) and transcriptional activators binding to these sequences modulate transcription. Finally, distal enhancer elements, which can be found far from the transcription initiation site in either direction and orientation, constitute another group of DNA targets for factors modulating pol II activity.

Several distinct groups of proteins interact with each segment of a typical class II nuclear gene promoter during pol II–catalyzed transcription. The general initiation factors TFIIB, -D, -E, -F, and -H must assemble on the core promoter with pol II before transcription begins. TFIID is the only one of

*Figure 1* Preinitiation complex assembly begins with TFIID recognizing the TATA element, followed by coordinated accretion of TFIIB, the nonphosphorylated form of pol II and TFIIF, TFIIE, and TFIIH. If ATP is available to support promoter melting, an open complex can form. In the presence of appropriate NTPs, initiation will occur. Promoter clearance follows phosphorylation of the C-terminal domain (CTD) of the large subunit of pol II. During promoter clearance, TFIIB, TFIIE, and TFIIH disengage from the polymerase, creating an elongation complex (phosphorylated pol II and TFIIF). Following termination, a phosphatase recycles pol II to its nonphosphorylated form, allowing the enzyme to reinitiate transcription. TBP (and TFIID) binding to the TATA box is an intrinsically slow step, yielding a long-lived protein-DNA complex. Efficient reinitiation of transcription can be achieved if recycled pol II reenters the preinitiation complex before TFIID dissociates from the core promoter. Coactivators (TFIIA and the TAFs) are shown in light gray. Reproduced with permission from Reference 122.



these components capable of sequence-specific DNA binding, and in the most general case the orderly process of transcription initiation begins with TFIID recognizing and binding tightly to the TATA box. Thereafter, the TFIID-TATA box complex directs accretion of the remaining general initiation factors and pol II to form the preinitiation complex or PIC, a large multiprotein-DNA assembly that supports accurate initiation of transcription at one or at most a few start sites (Figure 1; reviewed in 122). Thus the PIC is functionally equivalent to the much simpler *Escherichia coli* holoenzyme, which is composed of the core RNA polymerase subunits and one of a modest number of $\sigma$-factors (reviewed in 51).

Reconstitution of the pol II PIC in vitro has proved remarkably successful for mechanistic studies of basal transcription initiation. However, pol II–mediated transcription is considerably more complex in vivo. Several other transcription factors, both cellular and viral in origin, regulate the precise level of messenger RNA production from class II nuclear gene promoters (reviewed in 54). These proteins are often referred to as transcriptional activators. They modulate transcription by recognizing promoter proximal and/or distal enhancer DNA targets and participating in highly specific protein-protein interactions with components of the PIC and with each other. Efficiency of RNA production from pol II promoters depends, at least in part, on the half-life of the promoter-specific transcription complex and much effort is now being devoted to establishing good in vitro models of activator-dependent transcription initiation.

Work reported during this decade has extended significantly our understanding of the structural biology of messenger RNA transcription initiation in eukaryotes. X-ray crystallography has been used to examine protein-DNA complexes containing both core promoter elements (TATA and Inr). Moreover, X-ray and solution NMR methods have been applied successfully to studies of several transcriptional activators bound to their DNA targets. This review describes all published high-resolution structural studies of protein-DNA complexes pertaining to pol II transcription. In addition, modern perspectives on the problem of molecular recognition are presented.

## CORE PROMOTER BINDING FACTORS

Our high-resolution structural knowledge of core promoter binding factors includes examples of complexes with both TATA and Inr elements. The TATA box-binding protein recognizes its A/T-rich DNA target via minor groove interactions, inducing an unprecedented conformational change in the double helix. In contrast, YY1 binds to straight B-form DNA, presenting four "recognition helices" to the major groove.

### *TBP: Minor Groove Recognition and DNA Bending*

TFIID has been the focus of considerable biochemical and genetic study since its discovery in human cells in 1980 (86). DNA binding by human TFIID was first

demonstrated with the Adenovirus major late promoter (AdMLP) (128), revealing sequence-specific interactions between TFIID and the TATA element that are primarily mediated by its TATA box-binding protein (TBP) subunit. Publication of the sequence of yeast TBP by five laboratories was followed rapidly by the sequences of homologous genes from various eukaryotes and archaebacteria (amino acid identities within the conserved 180 residue C-terminal portion range between 38% and 100%; reviewed in 93). Recombinant TBP alone can bind both general and regulatory factors and direct PIC assembly and basal (core promoter) transcription in vitro (reviewed in 122). Activator-dependent transcription, however, requires TBP and the remaining subunits of TFIID, the TBP-associated factors or TAFs (reviewed in 10), and a number of non-TAF coactivators (reviewed in 37).

Following TFIID or TBP recognition of the TATA element in a typical class II nuclear gene promoter, TFIIB is the next general transcription factor to enter the PIC. The resulting TFIIB-TFIID(TBP)-DNA platform is in turn recognized by a complex of pol II and TFIIF (pol/F). Mutations in TFIIB alter pol II start sites in yeast, as do mutations in the large subunit of pol II, providing compelling evidence for its function as a precise spacer/bridge between TBP and pol II that determines the transcription start site (78, 113). Finally, PIC assembly is completed with recruitment of TFIIE and TFIIH. In the presence of nucleoside triphosphates, strand separation at the transcription start site occurs to give an open complex, the C-terminal domain of the large subunit of pol II is phosphorylated, and pol II initiates transcription and is released from the promoter. During elongation in vitro, TFIID can remain bound to the core promoter supporting rapid reinitiation of transcription by pol II and the other general factors (Figure 1; reviewed in 155). An abbreviated PIC assembly mechanism has also been proposed, following recent discoveries of various pol II holoenzymes containing many if not all of the general initiation factors other than TFIID (reviewed in 66).

Researchers have reported (*a*) structures of *Arabidopsis thaliana* TBP2 complexed with the AdMLP TATA element (5′TATAAAAG3′) (60, 61), (*b*) the C-terminal portion of yeast TBP complexed with the yeast CYC1-52 TATA element (5′TATATAAA3′) (62), and (*c*) the C-terminal portion of human TBP complexed with the AdMLP TATA element (95) (Figure 2). Although the three co-crystal structures differ slightly in detail, they suggest an induced-fit mechanism of protein-DNA recognition that has been confirmed by more than a dozen high-resolution co-crystal structures of wild-type *A. thaliana* TBP2 complexed with TATA element variants (Patikoglou et al, to be published). DNA binding is mediated by the protein's curved, eight-stranded, antiparallel $\beta$-sheet, which provides a large concave surface for minor groove and phosphoribose contacts with the eight base-pair TATA element. The 5′ end of standard B-form DNA enters the underside of the molecular saddle, where the C-terminal

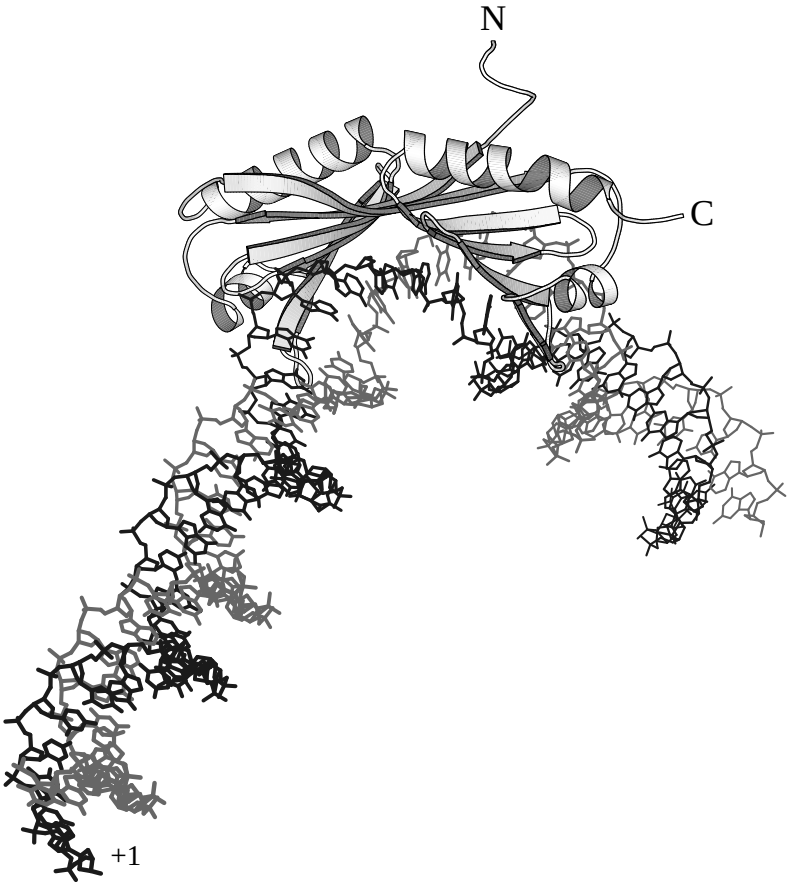



*Figure 2* MOLSCRIPT (70) drawing of the structure of TBP2 from *Arabidopsis thaliana* complexed with the AdMLP TATA element (61). The molecular saddle (N- and C-termini-labeled) is depicted with a ribbon drawing and the DNA is shown as a stick figure with the transcription start site labeled with +1. Adapted from Reference 61.

portion of TBP produces an abrupt transition to an unprecedented, partially unwound form of the right-handed double helix induced by insertion of two phenylalanine residues into the first TpA base step (roll angle = 40–52°, rise distance = 5.4–5.9 Å). Thereafter, the widened minor groove face of the unwound, smoothly bent DNA is approximated to the underside of the saddle-shaped protein, burying a total surface area of about 3,100 $Å^2$, permitting direct interactions between protein side chains and the minor groove edges of the central six base pairs. A second large kink is induced by insertion of two

phenylalanine residues into the base step between the last two base pairs of the TATA element, and there is a corresponding abrupt return to B-form DNA (roll angle = 40–45°, rise distance = 5.3–5.7 Å).

Despite this massive distortion, Watson-Crick base pairing is preserved throughout and there appears to be no helical strain induced in the DNA, because partial unwinding is compensated for by right-handed supercoiling of the double helix. Side chain–base contacts are restricted to the minor groove, including the four phenylalanines described above plus five hydrogen bonds and a large number of van der Waals contacts. There are no water molecules mediating side chain–base interactions, and the majority of the hydrogen bond donors and acceptors on the minor groove edges of the bases remain unsatisfied (13 out of 17 in the AdMLP TATA box). Detailed analysis of the TBP2-DNA co-crystal structure at 1.9 Å resolution demonstrates that the protein also undergoes a modest conformational change on DNA binding, involving a twisting motion of one domain with respect to the other (60).

DNA packaging into nucleosomes involves wrapping a double helix around the histone octamer (radius of curvature = 45 Å). 5′TA3′-containing sequences and certain trinucleotide sequences (5′AAA3′ and 5′AAT3′) show strong preferences for approximating the minor groove to the basic surface of the histone octamer (24, 124). Sequence-dependent nucleosome positioning correlates with bending A+T-rich sequences toward the minor groove, and packaging of TATA elements into nucleosomes probably results in minor groove compression, precluding TBP binding. Conversely, a preformed PIC remains transcriptionally active after nucleosome assembly (153), and recombinant yeast TBP alone prevents nucleosome-mediated repression of transcription (88). Thus, minor groove recognition and bending of DNA by TBP may provide a mechanism for stabilizing the PIC through multiple rounds of transcription initiation in the presence of chromatin (114).

TBP's unusual mode of DNA binding and recognition may also explain the profound toxicity of minor groove-binding anti-tumor agents, such as netropsin and distamycin A. These drugs interact as monomers in the minor groove of B-form DNA, displacing a spine of water molecules, donating hydrogen bonds to A+T-rich sequences, and narrowing the minor groove (19, 68). In at least one case, two molecules of distamycin A have been shown to bind side by side in the minor groove of B-form DNA (12). These two similar modes of drug binding appear opposite to the way TBP recognizes the TATA element, and distamycin A is a potent inhibitor of TBP-DNA complex formation in vitro (14). Other relevant work on TBP includes

- examination of the rather complicated kinetics of TATA element binding (second-order rate constant for simultaneous binding and bending is about

$10^6 M^{-1}s^{-1}$; see 53, 106) and dissociation (half-life is about two hours in vitro; see 53);

- TBP binding enhanced by prebending of the DNA toward the major groove and vice versa (107); and
- DNA deformation extending outside the confines of the TATA box during TBP binding (142).

Crystal structures of a TFIIB-TBP-TATA element ternary complex (94) and two TFIIA-TBP-TATA element ternary complexes (38, 143) have also been determined, enabling us to build a plausible model of the TFIIA-TFIIB-TBP-TATA element quaternary complex (Figure 3). Both TFIIA and TFIIB recognize the preformed TBP-DNA complex by interacting with the N- and C-terminal stirrups, respectively, of the molecular saddle, and with the phosphoribose backbone of the deformed double helix.

### *SRY/LEF-1: Two Other Minor Groove Binding Proteins*

DNA deformation by TBP may also be important for coordinating and/or stabilizing PIC assembly and activator-PIC interactions. PIC assembly around a bend could produce a more compact complex on the promoter. In addition, DNA bending by TBP could aid in the looping of DNA to bring remotely bound transcription factors closer to the core promoter for interactions with components of the PIC. In this sense, it would function similarly to the high-mobility-group (HMG) proteins (81, 149) and to the bacterial protein catabolite activator protein (CAP) (129), which are thought to activate transcription through bending of DNA upstream of the initiation site (reviewed in 144).

NMR solution studies of the HMG domains of SRY (149) and LEF-1 (81) have revealed similar $\alpha$-helical proteins, which both present L-shaped, concave binding surfaces to the DNA. Although these proteins are entirely unlike TBP, they too demonstrate DNA bending and unwinding that is mediated by intercalation of a hydrophobic amino acid into a base step from the minor groove face of the double helix. In LEF-1 a methionine side chain is inserted between two A:T base pairs, and in SRY an isoleucine is inserted into an ApA base step. Like TBP, both SRY and LEF-1 make additional side chain–minor groove base edge interactions that are primarily hydrophobic, further distorting the trajectory of the double helix axis. In addition to the minor groove interactions, basic residues from the C-terminus of the LEF-1 HMG domain bind to the phosphoribose backbone on the major groove face of the DNA.

### *YY1: Initiator Element Recognition*

YY1 is a human GLI-Kruppel-related protein capable of transcriptional activation or repression, and sequence-specific Initiator (Inr) element binding

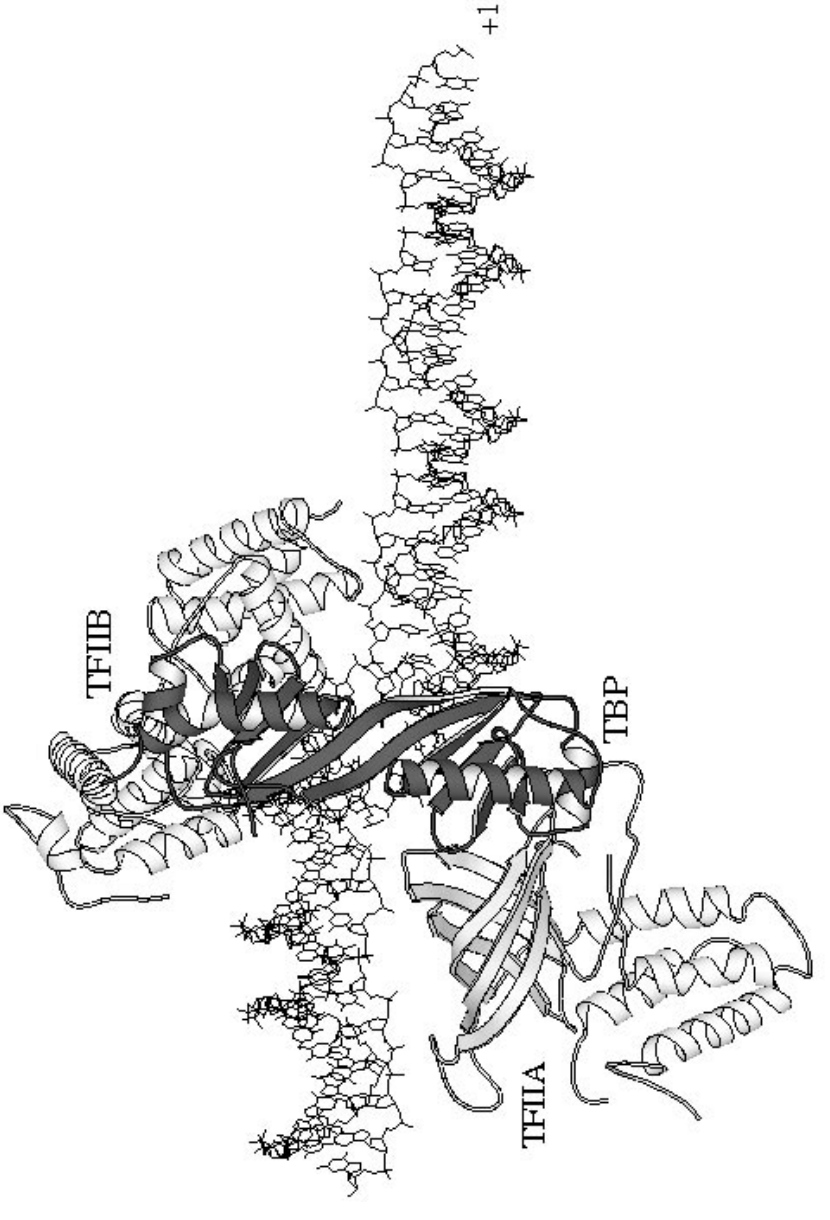



*Figure 3* Drawing of a model of the TFIIA-TFIIB-TBP-DNA quaternary complex derived from structures of the TFIIB-TBP-DNA (94) and TFIIA-TBP-DNA (38, 143) ternary complexes.

(reviewed in 44). In vitro studies have demonstrated that YY1 recognition of the Inr element of the Adeno-associated virus (AAV) P5 promoter supports specific, unidirectional transcription initiation from two adjacent start sites (134; see 40 for review). Despite the presence of a TATA box in the P5 promoter, this process is independent of TBP (145). YY1-directed transcription initiation from the P5 promoter begins with assembly of a YY1-DNA complex that recruits TFIIB, creating a YY1-TFIIB-DNA complex that in turn brings pol II to the promoter (146). Specific, unidirectional YY1-dependent initiation of transcription can proceed in the presence of nucleotide triphosphates if the P5 template is negatively supercoiled (145). This situation is precisely analogous to TBP-directed transcription initiation by TBP, TFIIB, and pol II in the absence of other transcription factors from a negatively supercoiled template containing a TATA element (108).

The co-crystal structure of the DNA-binding domain of human YY1 recognizing the AAV P5 promoter Inr element has been determined at 2.5 Å resolution (55). The structure and the results of complementary biochemical studies (55, 146) permit detailed functional analyses of protein-DNA interactions, providing further insights into YY1-directed, TBP-independent transcription initiation. The YY1-Inr element complex demonstrates many features of DNA recognition by zinc-finger proteins (see below). Three of YY1's four zinc fingers make multiple contacts with the major groove edges of bases (finger 1 makes a single base contact), and the observed side chain–base contacts explain the YY1-binding consensus. In the absence of any knowledge of the structure of the YY1-Inr element complex, it would appear that YY1 might inhibit rather than promote transcription from the P5 promoter by rendering the transcription start sites inaccessible to the active center of pol II. However, the structure shows that YY1 interacts with both DNA strands upstream of the two base pairs where mRNA synthesis is initiated but interacts almost exclusively with the template strand downstream of these two bases (Figure 4). In fact, downstream of the two start sites the non-template strand is almost completely exposed to solvent and is presumably accessible to the active center of pol II.

On a supercoiled P5 promoter template, therefore, YY1, TFIIB, and pol II could together form a stable complex capable of supporting strand separation beginning at the two start sites and extending only in the 3′ direction. Transcription initiation events occurring either upstream or in the opposite direction would almost certainly lead to dissociation of YY1 from the Inr element. Compelling, albeit indirect, support for this assertion comes from the results of studies of the AAV P5 promoter containing bubble mismatches (146). YY1 binding and accurate YY1-directed transcription initiation occur without supercoiling, if the bubble mismatch preserves the template strand and does not extend upstream of the YY1-dependent transcription start sites. Presumably, energy from negative supercoiling is required in the absence of a preformed

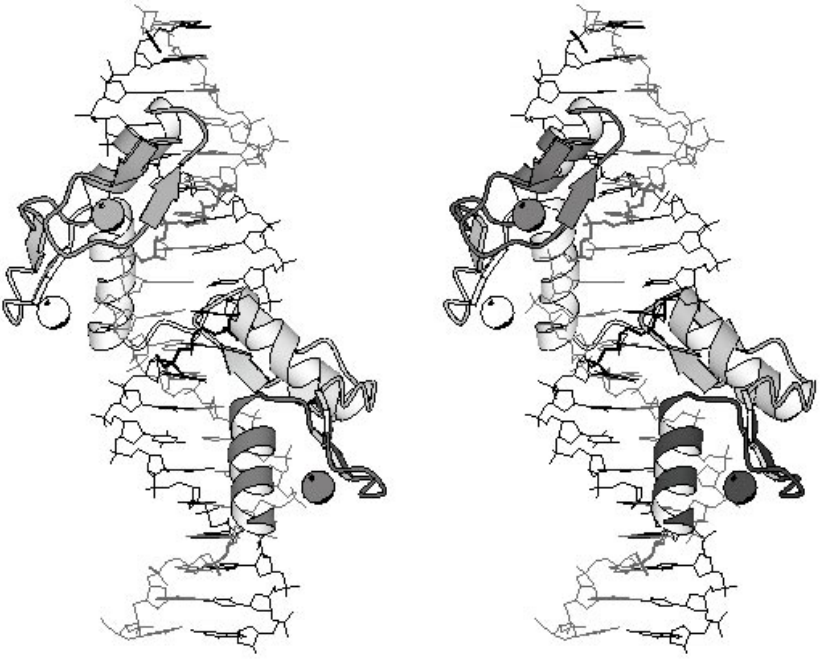


*Figure 4* Stereodrawing of the YY1-DNA complex (55). The protein is shown as a ribbon representation, the DNA as a stick model, and the zinc ions as spheres. The template strand of the DNA is gray, and the non-template strand is black. RNA synthesis proceeds downward in this view of the complex. Adapted from Reference 55.

bubble to facilitate DNA strand separation. The proposed model is further substantiated by the results of gel mobility shift studies that demonstrate specific binding of YY1 to an oligonucleotide in which the part of DNA that corresponds to the nascent transcript had been replaced with RNA (55). The complex between ΔYY1 and the DNA/RNA hybrid oligonucleotide represents an analog of the ternary DNA/RNA hybrid-protein complex formed immediately after initiation of transcription has occurred. YY1 could, therefore, remain bound to the template not only during the DNA strand separation but also after the nascent mRNA transcript has been synthesized.

## PROMOTER PROXIMAL/DISTAL ENHANCER BINDING FACTORS

Transcription initiation of a class II nuclear gene in response to developmental and/or environmental signals is controlled by regulating the assembly of a large multiprotein complex on the gene's promoter. In their simplest form,

interactions regulating pol II activity involve direct protein-protein contacts between components of the general transcription machinery (TFIID, TFIIB, pol II, TFIIF, TFIIE, and TFIIH) and transcriptional activators (bound either to promoter proximal or distal enhancer elements). Indirect interactions between the basal machinery and transcriptional activators mediated by non-TAF coactivators, including TFIIA, have also been observed. Direct contact between an activator and a given basal factor could stabilize an intermediate in PIC assembly, ultimately leading to an increased rate of PIC formation or increased PIC stability. Conformational changes important for stronger binding at this or a later stage in the assembly, disassembly, or catalytic processes may also be involved.

In vivo footprinting of the promoter proximal regions of some liver-specific genes has demonstrated that many transcriptional activators appear to be bound simultaneously (121), which is consistent with the view that two or more activators can exert synergistic effects on transcription through concerted interactions with multiple components of the PIC (reviewed in 6, 13, 16, 37). In direct support of this hypothesis, Sauer et al (126, 127) have demonstrated that synergy between two different activators (Bicoid and Hunchback) bound to the same promoter results, at least in part, from specific interactions with two distinct TAFs that enhance TFIID recruitment.

Our high-resolution structural knowledge of the transcriptional activators is currently limited to their interactions with DNA. Since structures of two homeodomain-DNA complexes were published at the beginning of this decade (63, 102), there has been a dramatic acceleration in the pace with which protein-DNA complexes have been determined by both X-ray crystallography and NMR. The current database of high-resolution (better than 3 Å) structures of protein-DNA complexes includes more than 30 structures, which differ dramatically in size, shape, and mode of interaction with DNA. This review examines important aspects of each structural class of transcriptional activators, with reference to more specialized timely reviews, wherever possible.

## *Helix-Turn-Helix Proteins*

Until recently, the homeodomain proteins were the only eukaryotic transcription factors known to contain the helix-turn-helix (HTH) motif, typical of prokaryotic repressors (reviewed in 104). $\beta$-Turns found in canonical HTH proteins are four residues in length with glycine at the second position. Various structures examined below reveal a new subfamily of eukaryotic DNA-binding proteins known as HTH variants, for which the structural definition is loosened to permit loops instead of turns.

HOMEODOMAIN PROTEINS These proteins were the first eukaryotic transcription factors to be studied with high-resolution tools. Publication of the structures

of the Antennapedia homeodomain-DNA complex (102) and the Engrailed homeodomain-DNA complex (63) ushered in a new era in structural biology. The homeodomain is a compact 60-residue DNA-binding domain that consists of three $\alpha$-helices folded around a hydrophobic core and a flexible N-terminal arm that becomes ordered only on DNA binding. These structures, determined respectively in the Wuthrich and Pabo laboratories, confirmed earlier predictions that the eukaryotic homeodomain proteins would present their so-called recognition helices to the major groove of DNA, like their prokaryotic counterparts (Figure 5). However, the precise details of the observed protein-DNA interactions were significantly different from those seen in co-crystal structures of the dimeric repressor proteins (reviewed in 104, 140). Additional work published on canonical homeodomain-DNA complexes includes further solution NMR work on the Antennapedia homeodomain-DNA complex (9, 116), and a 2.7 Å resolution X-ray structure of the yeast MAT$\alpha$2 homeodomain-DNA complex (152).

More recent studies of homeodomain-DNA complexes have focused on two important behavioral features of these proteins. First, a number of these proteins can function as either homo- or heterodimers. Second, solution NMR measurements with the Antennapedia homeodomain-DNA complex demonstrated the presence of rapidly exchanging solvent molecules in the interface between the protein and the major groove face of the double helix (117). Co-crystal structures of the Paired homeodomain homodimer (151) and of the Mat$\alpha$1/Mat$\alpha$2 homeodomain heterodimer (77) have been determined at 2.0 and 2.5 Å, respectively. Both systems show considerable DNA bending, cooperative protein-protein interactions, and contacts with an extended recognition site that exceeds the typical homeodomain binding site (5′TAAT3′) (see 75 for a review of the thermodynamic consequences of cooperative interactions between two DNA-binding domains).

These dimer structures, and the structure of the Even-skipped homeodomain-DNA complex (52), were determined at sufficiently high resolution to identify a significant number of water-mediated protein-DNA interactions. Water-mediated contacts are particularly extensive in the paired homeodomain homodimer complex (18 water molecules per homeodomain-DNA interface), which is reminiscent of the trp repressor-operator complex (103). When the trp repressor-operator complex structure was first published, there were suggestions that the complex did not include a biologically relevant binding site (139). In addition, there was widespread resistance to the premise that water molecules could act as specific bridges within any protein–nucleic acid interface. Recent X-ray and biophysical studies of the trp repressor-operator complex have resolved the specific dispute concerning the choice of crystallization DNA (46, 58, 74, 135). More generally, the concept of water bridges in specific macromolecular complexes has gained acceptance.

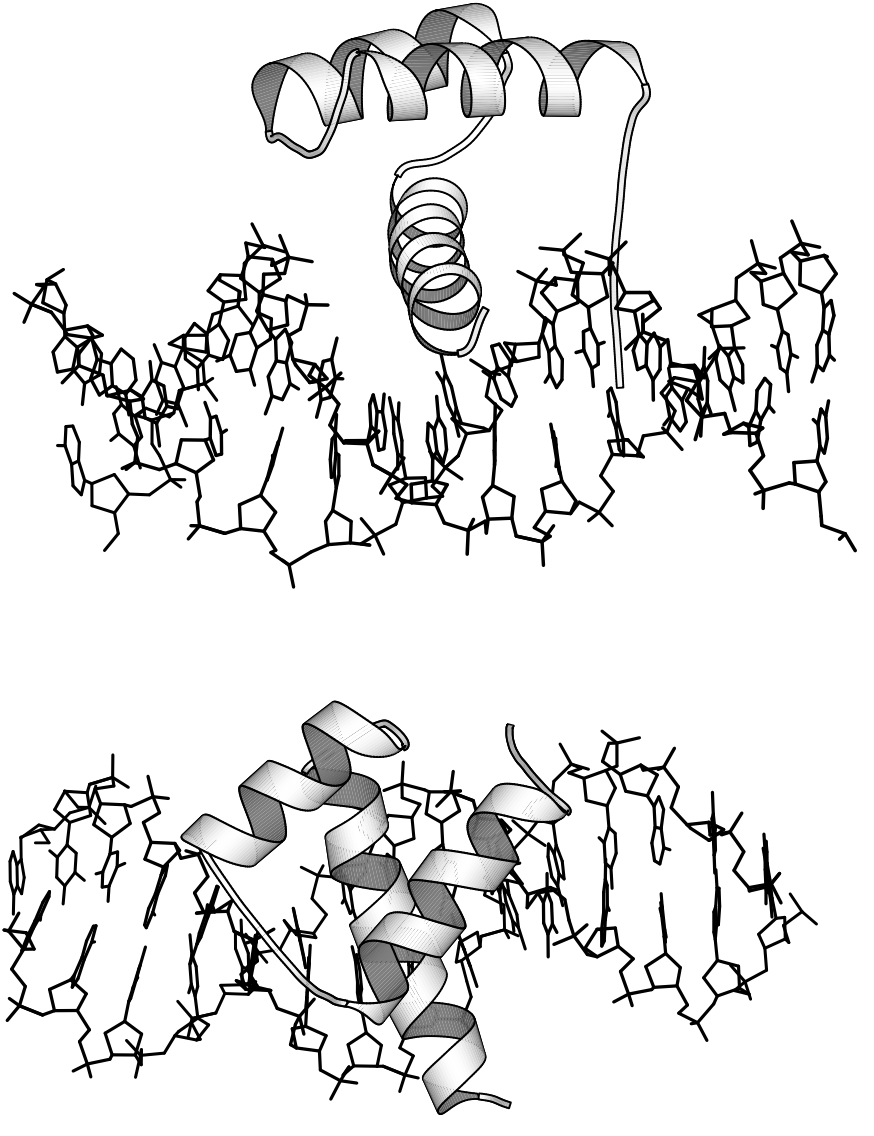


*Figure 5* Drawing of the Engrailed homeodomain-DNA complex (63). The protein is shown as a ribbon representation, with DNA as a stick figure. Profile (*right*) and face (*left*) views show presentation of the recognition helix to the major groove, with additional interactions between the N-terminus and the adjacent minor groove. Adapted from Reference 63.

MYB This mammalian transcriptional regulator consists of three imperfect direct repeats, the second and third of which are responsible for sequence-specific DNA binding (reviewed in 98). As predicted from amino acid sequence comparisons, the third repeat of c-Myb includes three $\alpha$-helices folded around a tryptophan-rich hydrophobic core and resembles the HTH homeodomain (98). c-Myb, however, is an HTH variant because the turn in this protein is one amino acid longer than the usual four and has the sequence Leu-Pro-Gly-Arg-Thr. The NMR solution structure of the second and third repeats of c-Myb complexed with DNA reveal another variation on the homeodomain theme (99). The recognition helices of repeats two and three are both presented to the major groove and are aligned C- to N-terminus, providing an extended DNA-contact surface that spans eight base pairs. This arrangement has not been observed for any other HTH protein or variant, but does resemble the arrangement of recognition helices in zinc-finger protein–DNA complexes (29, 55, 110, 111) (see below).

WINGED-HELIX PROTEINS These proteins represent another biologically important group of HTH variants. Hepatocyte nuclear factor-3$\gamma$ (HNF-3$\gamma$) is a mammalian tissue-specific transcription factor, containing a highly conserved 110-residue DNA-binding domain, shared by a large gene family of transcription factors found in various organisms (reviewed in 72). The structure of the HNF-3$\gamma$–DNA complex was determined at 2.5 Å resolution (17). The DNA-binding domain of HNF-3$\gamma$ is a winged helix $\alpha/\beta$ HTH variant that is topologically identical to the globular portion of histone H5 (118) (Figure 6). Like the homeodomains, HNF-3$\gamma$ binds as a monomer to a slightly bent, B-form DNA with the recognition helix lying in the major groove and other regions of the protein making a variety of base and DNA backbone contacts, some of which are water mediated.

CAP-LIKE DOMAINS These domains are the DNA-binding portions of various eukaryotic transcriptional activators, including heat shock factor (21, 49) and Ets-containing proteins (23, 79, 150). (For an extremely detailed review of CAP structure and function, see 140.) Unlike CAP, the Ets domain of the transcription factor PU.1 binds DNA as a monomer, presenting its recognition helix to the major groove and inducing a modest, smooth bend in the double helix (65). Like other eukaryotic HTH variants, the Ets domain-DNA complex is stabilized by water-mediated interactions and by other regions of the protein, making a variety of base and DNA backbone contacts.

POU DOMAINS These proteins contain two discrete globular portions, including a POU homeodomain and a POU-specific domain, that are tethered to one another by a flexible linker (reviewed in 123). Two NMR solution structures of

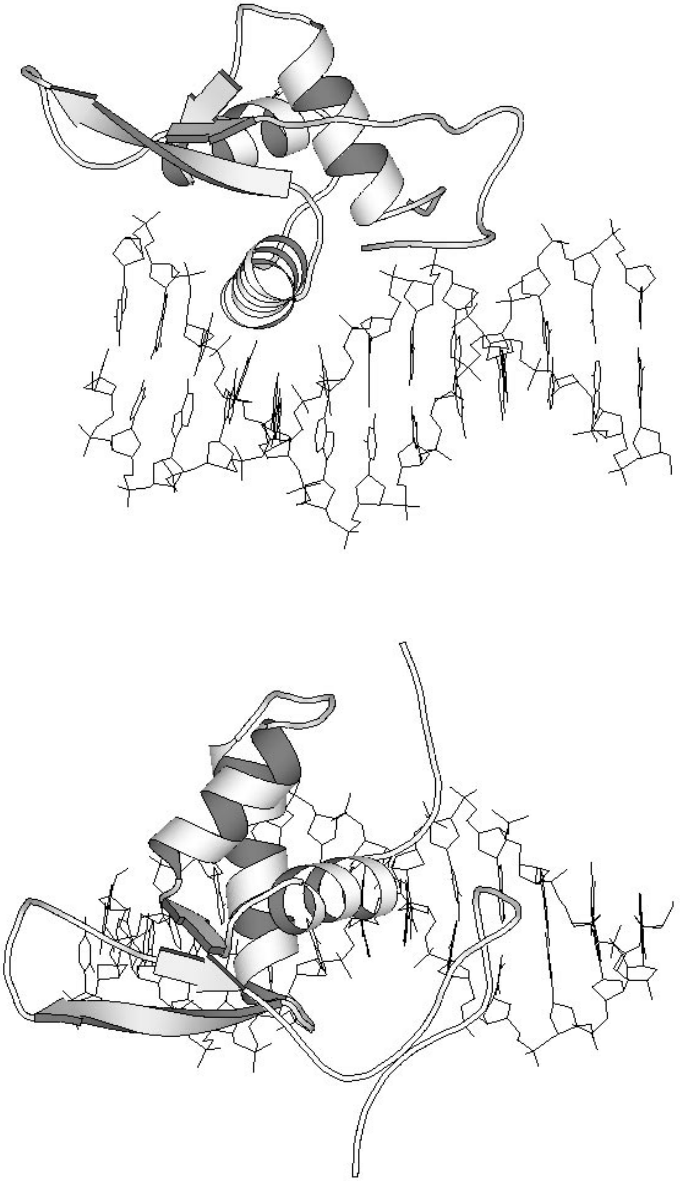


*Figure 6* Drawing of the HNF3γ-DNA complex (17), using views comparable to those in Figure 5. The winged-helix motif binds DNA by presenting the recognition helix to the major groove, with two wing-like loops interacting with flanking portions of the phosphoribose backbone and the adjacent minor groove. Adapted from Reference 17.

the POU-specific domain of the mammalian transcription factor Oct-1 demonstrate a striking similarity to the bacteriophage λ repressor DNA-binding domain (3, 22). Overlays of the POU-specific domain NMR structure on those of bona fide HTH proteins revealed a longer turn (Tyr-Gly-Asn-Asp-Phe-Ser) and an extension of the C-terminus of the first of the HTH helices. Although the POU homeodomain and the POU-specific domain were known to bind to adjacent recognition sites, it was not until Pabo and coworkers reported the crystal structure of the Oct-1 POU domain at 3 Å resolution that the molecular details of POU domain function were fully understood (64). This structure revealed two well-defined globular domains, each presenting a recognition helix to the major groove, connected by a flexible linker that is not visible in the electron density map (Figure 7). There are no interdomain interactions, and cooperative binding by the two domains is mediated by overlapping phosphoribose contacts near the center of the POU domain–binding site. Like the dimeric homeodomain

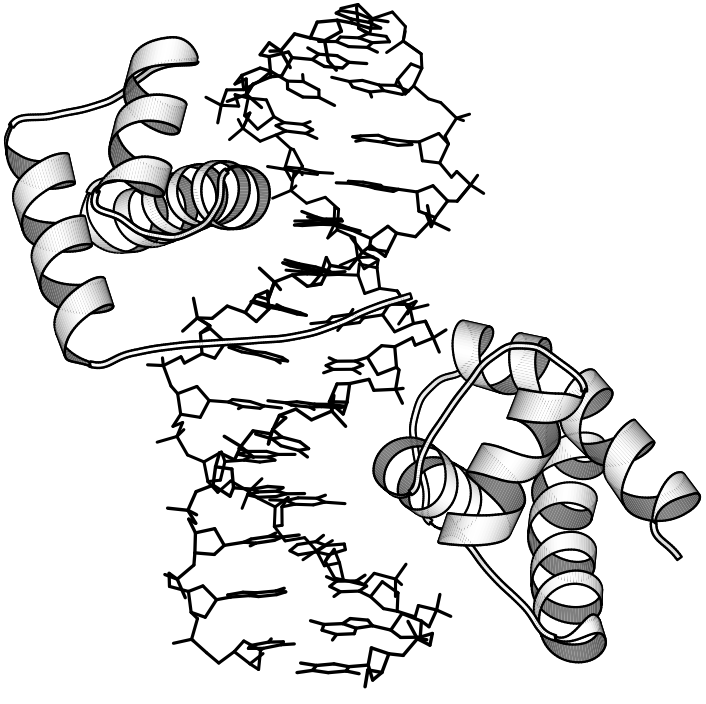


*Figure 7* Drawing of the POU domain-DNA complex (64), using a view comparable to the right panels of Figures 5 and 6. The linker connecting the two globular domains is not visible in the electron density map. Adapted from Reference 64.

structures, covalent tethering of two globular subdomains allows the POU domain to recognize a DNA sequence that is considerably longer than the four base pairs normally bound by a single homeodomain (see 57 for a detailed thermodynamic analysis of the consequences of tethering two DNA-binding domains together).

PAIRED DOMAINS These domains are found in the Pax proteins, which play important roles in *Drosophila* development (reviewed in 42). The crystal structure of a Paired domain–DNA complex, obtained at 2.5 Å resolution, reveals two globular subdomains that both resemble the homeodomain (154). Unlike the POU domain, only the N-terminal subdomain presents its recognition helix to the major groove. The C-terminal subdomain makes backbone contacts with two adjacent segments of DNA in the crystal.

## *Fibrous Proteins*

In contrast to the relatively compact, globular HTH proteins, there is another large family of transcriptional activators that resemble coiled-coil or fibrous proteins such as collagen (coiled-coil proteins are reviewed in 18). The coiled-coil transcription factors function as homo- and/or heterodimers, using their $\alpha$-helical basic regions to grip the double helix in a scissors-like fashion, making sequence-specific side chain–base contacts in the major groove. McKnight and coworkers (73, 147) anticipated the architecture of this class of transcriptional activators and set the stage for the intellectual synergy between structural and molecular biologists that followed (reviewed in 25, 30).

BASIC REGION/LEUCINE ZIPPER PROTEINS McKnight's landmark papers focused on this class of proteins. Co-crystal structures of homodimers of the basic region/leucine zipper portion of the yeast transcription factor GCN4 bound to two different recognition sites (27, 67) and X-ray (97) and NMR solution (125) studies of dimeric leucine zipper portions confirmed the scissors grip model (Figure 8). Dimerization is mediated by the leucine zipper portion of the protein, which contains a characteristic leucine heptad repeat that pairs to form a canonical left-handed, $\alpha$-helical coiled-coil (20). Functional studies using circular dichroism spectroscopy documented that the basic region undergoes a random coil–to–$\alpha$-helix folding transition when it binds to its cognate DNA (148). Thus DNA sequence recognition by the basic region/leucine zipper transcription factors represents another example of induced fit, albeit primarily at the protein level. The two GCN4 co-crystal structures differ at the level of DNA structure. The double helix is linear for the 5′ATGACTCAT3′ AP-1 site and moderately bent for the 5′ATGACGTCAT3′ ATF/CREB site. These two sites consist of symmetrically oriented pairs of 5′ATGA3′ half sites separated by either one or two base pairs. The ATF/CREB site bend, accompanied

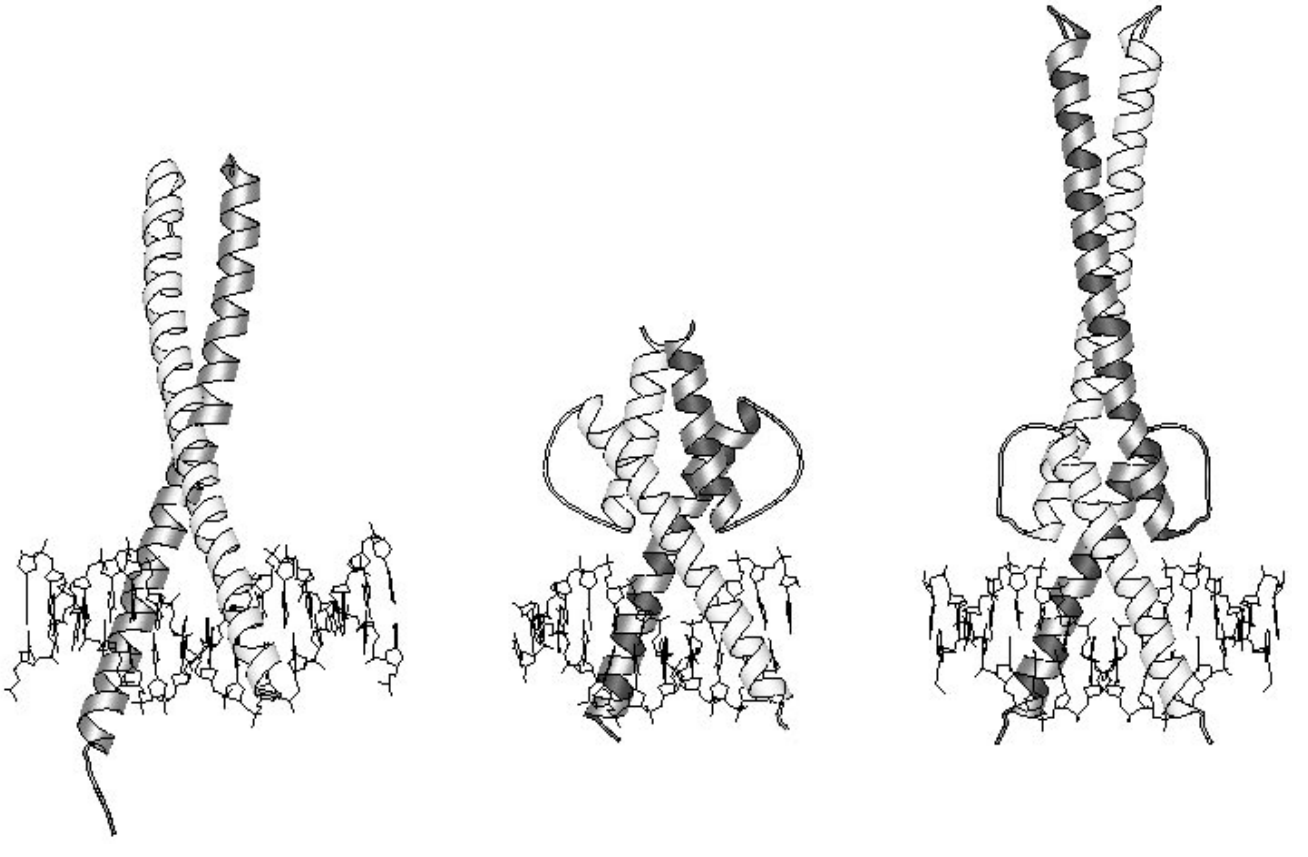


*Figure 8* Drawing comparing from left to right the structures of the basic region/leucine zipper protein GCN4-(27), the basic region/helix-loop-helix protein E47-(26), and the basic region/helix-loop-helix/leucine zipper protein Max-DNA (32) complexes. The right half of each homodimer is denoted with shaded ribbons. Adapted from References 26, 27, and 32.

by local unwinding, reduces the half-site spacing, permitting the basic region $\alpha$-helices of GCN4 to recognize each half site via the same side chain–base contacts seen in the complex of GCN4 with the AP-1 site. DNA-bending studies have demonstrated that the ATF/CREB site is intrinsically bent in the absence of GCN4, revealing that GCN4 can recognize both straight and curved B-form DNA (105). Schepartz and coworkers also demonstrated that GCN4 can induce bends in otherwise straight DNA. More recent progress on structural studies of these proteins includes the co-crystal structure of a fos/jun heterodimer bound to DNA (41).

BASIC REGION/HELIX-LOOP-HELIX/LEUCINE ZIPPER PROTEINS These proteins were also anticipated by molecular biologists well before structuralists prevailed. The basic region/helix-loop-helix/leucine zipper proteins are characterized by a highly conserved 60–100 residue motif comprised of two amphipathic $\alpha$-helices separated by a loop of variable length, as well as by their amino acid composition and sequence. These motifs dimerize by approximating the conserved hydrophobic faces of the $\alpha$-helices, forming a left-handed, parallel, four-helix bundle (Figure 8).

Baltimore and coworkers (91) first described conservation of this feature in several eukaryotic transcription factors and implicated it in dimerization and DNA binding. Subsequent work has confirmed their prediction and demonstrated that the helix-loop-helix motif is primarily responsible for dimerization. Most of these proteins possess a highly conserved basic region immediately N-terminal to the first helix, which mediates high-affinity, sequence-specific DNA binding. In addition, the second helix of many of these transcription factors is extended beyond the C-terminus of the four-helix bundle, where a leucine heptad repeat or zipper forms a left-handed, coiled-coil that extends the dimer interface.

High-resolution structural studies of basic region/helix-loop-helix/leucine zipper proteins and their basic region/helix-loop-helix counterparts have been carried out using X-ray crystallography. The structure of the homodimer of the DNA-binding portion of Max complexed with 5′CACGTG3′ (32) was the first of these structures to be determined (Figure 9). Additional related structures include homodimeric protein-DNA complexes of another basic region/helix-loop-helix/leucine zipper protein known as upstream stimulatory factor (31), and two basic region/helix-loop-helix proteins, E47 (26) and MyoD (83) (Figure 8). Like the basic region/leucine zipper proteins, members of these two transcription factor subfamilies recognize DNA via an induced-fit mechanism, in which the basic region undergoes a random coil–to–$\alpha$-helix folding transition when it binds to its cognate DNA (2, 24, 31). Basic region/helix-loop-helix and basic region/helix-loop-helix/leucine zipper proteins are also

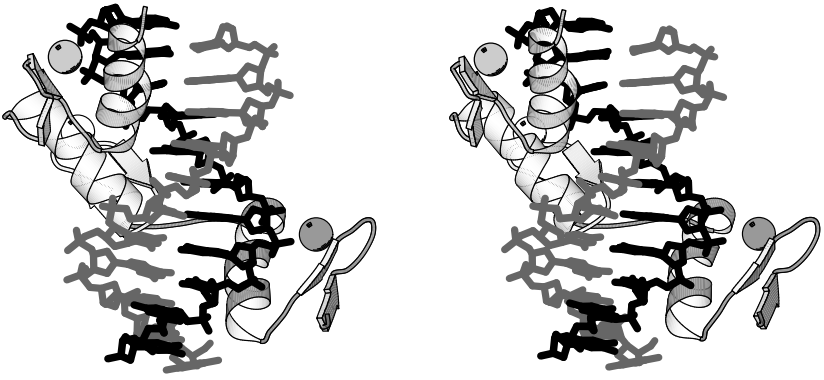


*Figure 9* Stereodrawing of the Zif268-DNA complex (110), using the same view as in Figure 4. The protein is shown as a ribbon representation, the DNA as a stick model, and the zinc ions as spheres. Adapted from Reference 110.

capable of forming heterodimers (within each subfamily), and considerable effort has gone into characterizing their homo- and heterodimerization properties (reviewed in 30). Although there are no published heterodimer structures, total chemical synthesis of a covalent Myc-Max heterodimer (11) suggests that it will eventually be possible to prepare heterodimer co-crystals suitable for X-ray diffraction studies.

## *Metallo-Proteins*

Many eukaryotic DNA-binding proteins contain zinc as an essential cofactor (reviewed in 8, 133). Only a subset of these proteins has been studied at high resolution using solution NMR and X-ray crystallography (see below). Structures of isolated zinc-containing domains derived from eukaryotic transcription factors include the N-terminal portion of TFIIB (156), the elongation factor transcription factor IIS (115), and a $Cys^3HisCys^4$ domain (7).

TRANSCRIPTION FACTOR IIIA This factor was the first gene-specific eukaryotic transcription to be described (28). Shortly after the gene for TFIIIA was cloned and sequenced, Klug and coworkers identified a repeated 27-residue motif that contains a zinc-binding $Cys^2His^2$ tetrad, now referred to as a zinc finger (89). Wright and coworkers determined the first structure of an isolated zinc-finger domain by solution NMR (76). Pavletich & Pabo determined the first co-crystal structure of a zinc-finger protein bound to DNA (110). The Zif268-DNA complex structure revealed three zinc fingers, each presenting their recognition helices to the major groove (Figure 9). Protein–nucleic acid contacts are limited to a single DNA strand, and each finger interacts with just

three successive base pairs. Despite the simplicity found in the Zif268 case, the co-crystal structures of Gli (five fingers; 111), Tramtrack (two fingers; 29), and YY1 (four fingers; 55) were considerably more complicated. Analyses of zinc-finger protein co-crystal structures revealed an underwound double-helix structure with a widened major groove intermediate between B and A forms (92). Physical studies with SP1 and Zif268 suggest that this unwound conformation is induced by zinc-finger protein binding (137). Additional biophysical analyses of zinc-finger proteins have documented that they can bind RNA/DNA hybrids, if Watson-Crick base pairing is maintained and the protein–nucleic acid contacts are restricted to the DNA strand (136).

STEROID/NUCLEAR RECEPTOR PROTEINS These proteins represent another important class of zinc-containing eukaryotic transcriptional activators, which are characterized by the presence of a $Cys^4$ double loop–zinc-helix DNA-binding motif (reviewed in 35). These proteins function as homo- and/or heterodimers, recognizing both inverted and directly repeated half sites with various spacings. To date, structures of five protein-DNA complexes have been published, including the glucocorticoid receptor DNA-binding domain homodimer complexed with an inverted pair of half sites three and four base pairs apart (Figure 10) (82), the estrogen receptor DNA-binding domain homodimer complexed with two distinct inverted pairs of half sites three base pairs apart (Figure 10) (130–132), and a heterodimer of the DNA-binding domains of the thyroid receptor and the 9-*cis* retinoic acid receptor complexed with a directly repeated pair of half sites four base pairs apart (119). The solution structures of monomeric glucocorticoid and estrogen receptor DNA-binding domains have also been determined by solution NMR (47, 133a). Comparisons of the structures of the homo- and heterodimers with one another and with the monomer demonstrate that dimerization is mediated by a variety of protein-protein interactions, which are made possible by folding transitions in the dimer interface. Additional structure determinations of other steroid/nuclear receptor DNA-binding domain homo- and heterodimers bound to response elements with different half-site orientations and spacings are under way (PB Sigler, personal communication).

GAL4 AND PPR1 These proteins are members of a large family of $Zn_2^{++}Cys_6$ transcription factors found in fungi (reviewed in 1). They homodimerize via a short $\alpha$-helical coiled-coil on DNA elements containing two half sites. To date, co-crystal structures of GAL4 (84) and PPR1 (85) have been obtained. The globular, zinc-containing portions of these two similar DNA-binding domains present $\alpha$-helices to the major groove, where side chain–base contacts are responsible for 5′CGG3′ half-site recognition. Despite these similarities, GAL4 and PPR1 recognize DNA elements with unequal half-site spacings (11

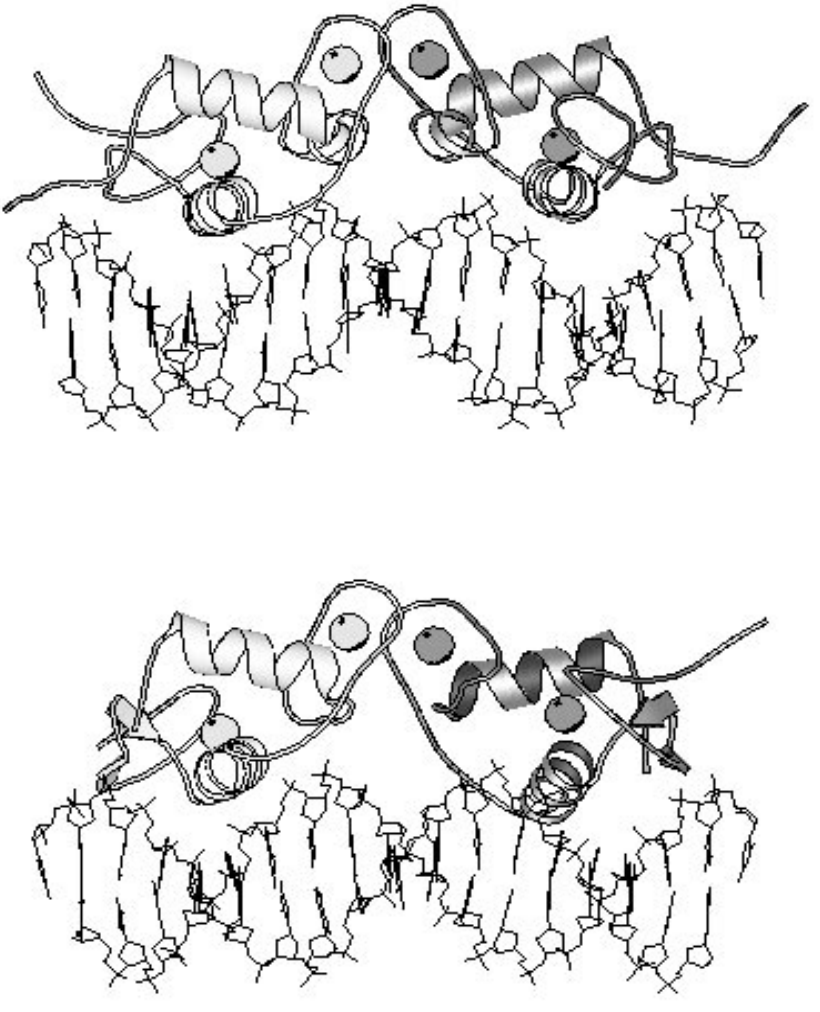


*Figure 10* Drawing comparing the structures of the glucocorticoid (*upper*; 82) and estrogen (*lower*; 130) receptor DNA-binding domains complexed with their respective recognition elements. The protein is shown as a ribbon representation (right half of each homodimer is denoted with a shaded ribbon), the DNA as a stick model, and the zinc ions as spheres. Adapted from References 82 and 130.

vs 6 base pairs), which are accommodated by markedly different interactions across the two homodimer interfaces.

GATA-1 This protein is an erythroid-specific transcription factor that regulates pol II activity within the erythroid cell lineage by binding as a monomer to the asymmetric consensus sequence 5′t/aGATAa/g3′ (reviewed in 101). Unlike the TFIIIA-like zinc-finger proteins and the GAL4 family members described above, the GATA-1 DNA-binding domain is a small $Cys^4$ $Zn^{++}$-containing $\alpha/\beta$ motif, which can function alone. The NMR solution structure of a GATA-1-DNA complex (100) reveals an $\alpha$-helix in the major groove of the recognition element. In addition, random coil regions participate in interactions with both the major and minor grooves.

P53 This tumor suppressor protein has been implicated in about half of all cases of human cancer (reviewed in 48). The co-crystal structure of the DNA-binding domain of p53 bound to its recognition element revealed important functional differences between wild-type and mutant p53 (15) (Figure 11). The overall fold of the DNA-binding domain had not been seen before, but its architectural features are by no means novel. Two antiparallel $\beta$-sheets pack face-to-face across an extended hydrophobic core. At one end of the core, two sheets form a barrel-like structure. The opposite face of the core is responsible for DNA binding, and contains two long loops, a two-stranded $\beta$-structure packed against an $\alpha$-helix, and the metal-binding region. An $\alpha$-helix and a loop are presented to the major groove and make side chain–base interactions, and an arginine side chain from another loop projects into the minor groove, making phosphoribose contacts. Of the six mutational hot spots, only two involve amino acid side chains that interact with DNA. The locations of the remaining four hot spots underscore the vulnerability of the p53 fold. The residues in question stabilize the structure of p53's DNA-binding edifice but make no direct or water-mediated contacts with nucleic acid. Mutations at any of these positions to any other residue would either denature p53 or reduce its thermal stability. In aggregate, p53 mutations at these four positions account for about 10% of all human cancers.

## *Miscellaneous Proteins*

E2 These papillomavirus proteins play central roles in regulating both transcription and viral DNA replication (reviewed in 45). The conserved 85-residue, DNA-binding domain of E2 binds with high affinity as an extremely stable homodimer to specific DNA sites. Work in the Sigler laboratory demonstrated that the E2 DNA-binding domain does not resemble any of the structures described above (50). E2 forms a dimeric eight-stranded antiparallel $\beta$-barrel structure that presents, on the surface of the barrel, two symmetrically disposed

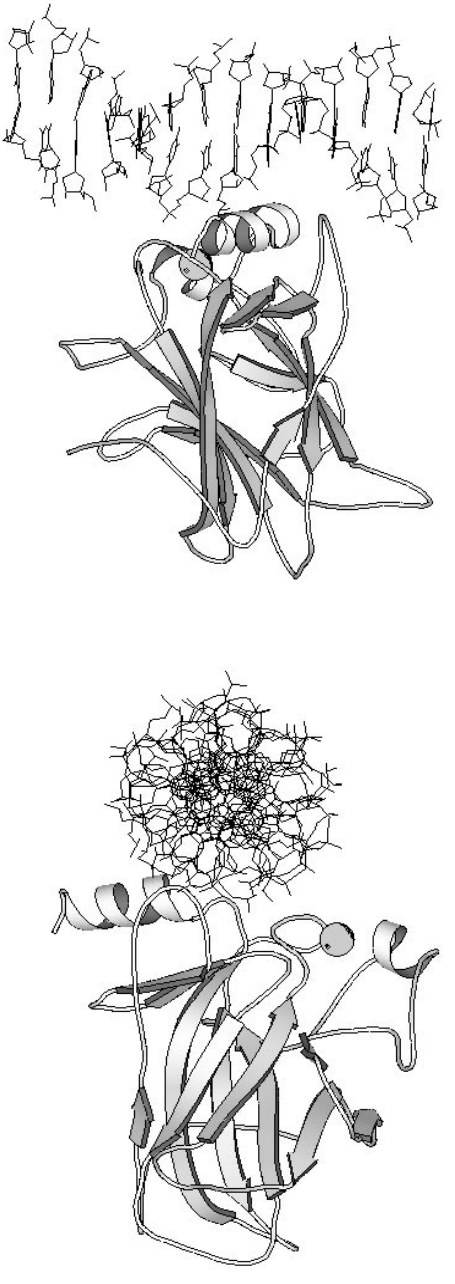


*Figure 11* Drawing of the p53-DNA complex (15), viewed in profile (*right*) and down the double helix axis (*left*). The protein is shown as a ribbon representation, the DNA as a stick model, and the zinc ion as a sphere. Adapted from Reference 15.

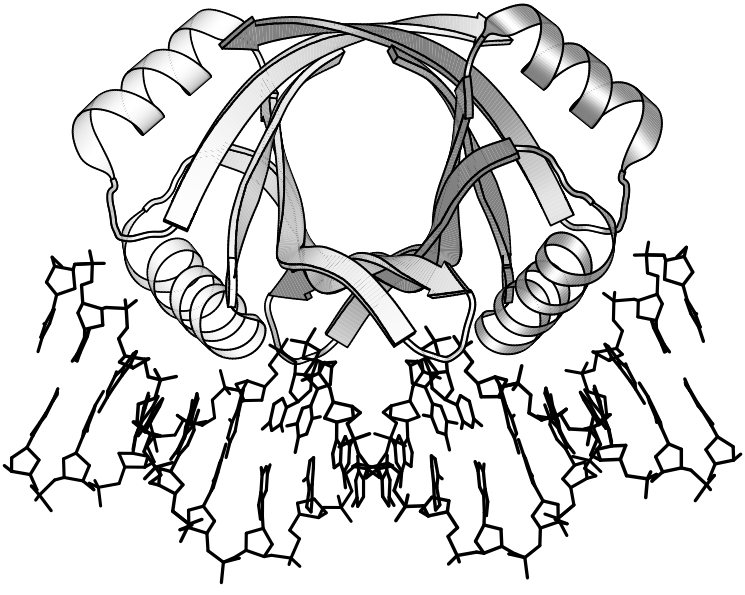


*Figure 12* Drawing of the E2-DNA complex (50). Each half of the homodimer is depicted with clear and shaded ribbons, respectively. Adapted from Reference 50.

recognition helices (Figure 12). Although $\beta$ barrels are a common architectural feature in proteins, E2 is the first structure in which a barrel is formed by the dimerization of two polypeptide chains. This feature explains why the dimer is so exceptionally stable (it cannot be dissociated, except under denaturing conditions). In addition to interactions stabilizing its seams, the interior of the barrel constitutes the hydrophobic core of the dimer, and two monomers would have strong incentive to mate and exclude water. The two recognition helices are positioned such that both cannot make simultaneous contact with linear B-form DNA. Instead, the DNA bends more or less smoothly around the E2 protein into an arc that engages both $\alpha$-helices with adjacent major grooves. All side chains that make direct contact with the target sequence are presented by the recognition helix. It is remarkable that the side chain–base interactions are interwoven: Protein side chains make contact with more than one base pair, and these base pairs in turn interact with more than one side chain.

REL Members of this class share a 300–amino acid Rel homology region, responsible for both DNA binding and homo- and/or heterodimerization (reviewed in 4). These proteins differ from most transcriptionally active, DNA-binding

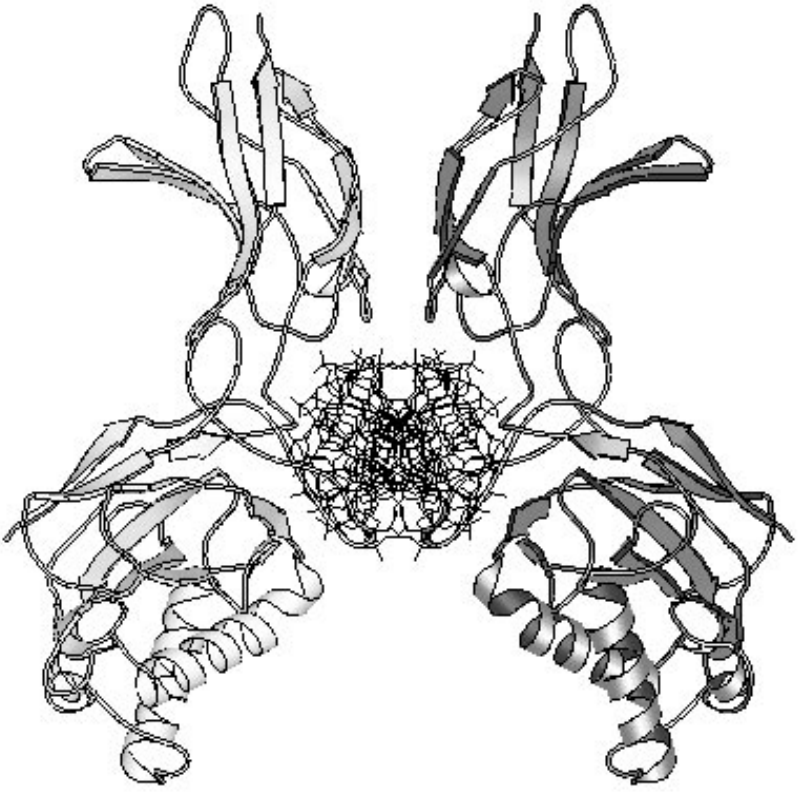


*Figure 13* Drawing of the NF$\kappa$B p50 homodimer-DNA complex (90, 39), viewed down the double helix axis. Adapted from Reference 90.

proteins. They do not possess a small subdomain or module capable of interacting with DNA; the complete Rel homology region is required for DNA recognition. Sigler's (39) and Harrison's (90) laboratories determined two highly similar and entirely novel three-dimensional structures of the NF-$\kappa$B p50 homodimer bound to distinct DNA targets. The three-dimensional structure of the symmetric p50 homodimer-DNA complex resembles a butterfly with the protein domains as wings surrounding a cylindrical B-form DNA (Figure 13). Each polypeptide chain folds into two domains, and there is no DNA-binding subdomain within the Rel homology region. Residues from the entire length of the protein contribute to the DNA-interaction surface.

Both domains contain $\beta$-sandwich core structures with immunoglobulin-like folds, which is a remarkable coincidence when one considers that NF-$\kappa$B

regulates transcription of genes encoding immunoglobulins. The larger N-terminal domain is connected to its C-terminal partner by a loop region that is well defined in the electron density map and contributes to interactions with the DNA. Dimerization is supported entirely by the smaller immunoglobulin-like C-terminal domain. The p50 homodimer embraces a slightly bent, slightly unwound, B-form double helix, donating a total of ten polypeptide loops for DNA contacts with both phosphoribose backbones and the major groove edges of some bases. The extensive protein–nucleic acid interaction surface (buried solvent-accessible area exceeds 3,500 $Å^2$) includes the entire $\kappa$B site, and each subunit makes contacts throughout the length of the recognition element. As in many of the protein-DNA complexes described above, both direct and water-mediated contacts are seen within the interface.

SERUM RESPONSE FACTOR This ubiquitous nuclear protein regulates cell proliferation and differentiation (reviewed in 96). Serum response factor contains a conserved DNA-binding region—a MADS box—found in many eukaryotic transcriptional activators. The co-crystal structure of the serum response factor MADS box bound to a serum response element (112) revealed a symmetric homodimer that forms an $\alpha$-$\beta$-$\alpha$ sandwich, which presents an antiparallel coiled-coil to the major grooves flanking a compressed minor groove in the center of a smoothly bent DNA element. Minor groove contacts are supported by the two N-termini that extend away from the helices bound in the major groove.

## PHYSICO-CHEMICAL INSIGHTS

The structural database of protein-DNA complexes furnished by X-ray crystallography, solution NMR spectroscopy, and other physical techniques is enormous. Compared with this bounty, our understanding of the problem of sequence-specific DNA recognition is lagging. The elegant simplicity of direct readout promised by early studies of bacteriophage repressor-operator complexes has not been realized. Moreover, the issue has become increasingly clouded by publication of a series of completely unanticipated co-crystal structures, beginning with the trp repressor-operator complex (103) and continuing to the present. As they emerged, some of the structures described above suggested that the only rule of importance governing protein–nucleic acid interactions was “No code for recognition” (87).

With many more structures now in hand, however, we have come to appreciate two universal features of sequence-specific DNA recognition. First, both transcription factors and DNA are hydrated macromolecules, and specific protein–nucleic acid complex formation requires rearrangement of water molecules in the vicinity of the contact surfaces. For TBP, the interaction

surfaces become completely dehydrated on complex formation. For the trp repressor, some homeodomains and a number of other DNA-binding proteins, a subset of the water molecules of hydration remains in the interface. Second, the proteins and/or the DNA frequently undergo conformational changes on complex formation. These unifying features may represent the key to understanding molecular recognition.

Both protein folding (141) and site-specific binding of proteins to DNA (43) are characterized by large negative heat capacity changes. Analysis of protein-DNA recognition, using ideas derived from protein folding, allowed Spolar & Record (138) to present a thermodynamic picture of the problem, which includes terms corresponding to the hydrophobic effect (i.e. burial of nonpolar surface area and extrusion of water molecules of hydration into the bulk solvent), loss of rigid body rotational and translational entropy, the polyelectrolyte effect (i.e. displacement of cations on binding), and loss of conformational flexibility (local folding or conformational changes and/or changes in vibrational properties coupled to binding). The large negative heat capacity change observed on specific protein-DNA complex formation originates from these entropic effects. In contrast, proteins incubated with noncognate DNA undergo no significant change in heat capacity, because they are thought to interact via electrostatic effects that leave their individual hydration properties largely unaffected.

The Spolar & Record analysis is briefly reproduced below, starting with the familiar change in the free-energy of association:

$$\Delta G_{assoc} = -RT \ln K_{assoc} = \Delta H_{assoc} - T\Delta S_{assoc}, \quad (1)$$

where $R$ is the universal gas constant, T represents temperature in degrees Kelvin, and $K_{assoc}$ represents the equilibrium association constant. What do these thermodynamic quantities mean in the context of protein-DNA complex formation? $\Delta H_{assoc}$, the enthalpy term, arises from the large number of short-range, noncovalent interactions between protein and DNA, including salt bridges, hydrogen bonds, weakly polar interactions, and van der Waals interactions. $T\Delta S_{assoc}$, the entropy term, results from changes in various factors when the complex forms, which are outlined in detail below. The favorable change in free-energy of association, $\Delta G_{assoc}$, for the TBP-DNA complex with an $\cong 10^9$ M equilibrium association constant, $K_{assoc}$, is $-10.5$ kcal mol$^{-1}$ at 25°C, corresponding to $\Delta H_{assoc} = 14.2$ kcal mol$^{-1}$ and $T\Delta S_{assoc} = 24.7$ kcal mol$^{-1}$ (M Brenowitz, unpublished observations). Thus the enthalpy term is unfavorable and is much less significant than the entropy term, suggesting that avoiding sterically and chemically unfavorable contacts in the protein-DNA interface is more important than making enthalpically favorable interactions.

Thermodynamic studies of specific protein-DNA associations have demonstrated that both the entropic ($T\Delta S_{assoc}$) and enthalpic ($\Delta H_{assoc}$) contributions

to $\Delta G_{assoc}$ vary with temperature, displaying large negative heat capacities ($\Delta C_{assoc}$) (43, 71). Rewriting Equation 1 using two characteristic temperatures, $T_H$ (where $\Delta H_{assoc} = 0$) and $T_S$ (where $\Delta S_{assoc} = 0$), gives

$$\Delta G_{assoc} = \Delta C_{assoc}[(T - T_H) - T \ln(T/T_S)], \quad (2)$$

where $\Delta H_{assoc} = \Delta C_{assoc}(T - T_H)$ and $\Delta S_{assoc} = \Delta C_{assoc} \ln(T/T_S)$ (5). Experimentally measured entropy changes due to specific protein-DNA complex formation ($\Delta S_{assoc}$) can be regarded as a sum of entropy changes due to the hydrophobic effect ($\Delta S_{HE}$), loss of rotational and translational freedom ($\Delta S_{rt}$), the polyelectrolyte effect ($\Delta S_{PE}$), and miscellaneous other effects ($\Delta S_{other}$) giving

$$\Delta S_{assoc}(T_S) = 0 = \Delta S_{HE}(T_S) + \Delta S_{rt} + \Delta S_{PE} + \Delta S_{other}. \quad (3)$$

The first three terms of the entropic summation are available from previous work on the protein-folding problem: $\Delta S_{HE}(T) = 0.32 \Delta A_{np} \ln(T/386)$, the entropic contribution due to the hydrophobic effect (5, 80, 141) ($\Delta A_{np}$ = nonpolar surface area buried during protein-DNA binding in $Å^2$); $\Delta S_{rt} = -50$ entropy units (cal $K^{-1}$ $mol^{-1}$) for a two-component system, estimated from studies of entropic changes arising from rigid body protein-protein association (33); $\Delta S_{PE}$ = the entropic contribution due to the polyelectrolyte effect (120). The fourth term in the summation, $\Delta S_{other}$, can be used to represent entropic changes arising from folding or conformational changes and/or from changes in vibrational properties that occur on specific—and only specific—binding.

Compelling experimental support for this thermodynamic analysis comes from the Sigler laboratory's work on the trp repressor-operator complex, the only system for which there is a full thermodynamic profile and high-resolution structures of protein alone, DNA alone, and the repressor-operator complex (71). Armed with a wealth of calorimetric and structural data, Sigler and colleagues were able to show that modulation of protein and/or DNA flexibility on specific complex formation contributes to the observed entropic and heat capacity changes. Additional, albeit indirect, support for the thermodynamic model of protein-DNA recognition presented above comes from the unusual co-crystal structures that stymied earlier attempts at developing a systematic understanding of the process. At one extreme, we find the basic regions of basic region/leucine zipper and basic region/helix-loop-helix/leucine zipper transcription factors undergoing specific, DNA-induced random coil–to–$\alpha$-helix folding transitions while binding to the major groove of B-form DNA. At the other extreme, we find the TATA box–binding proteins, which recognize the TATA element exclusively through minor groove determinants, inducing a modest conformational change in the protein and a massive DNA distortion involving unwinding, kinking, and bending of the double helix.

## CONCLUSION

During the past seven years, structural biologists have made tremendous contributions to our understanding of the mechanisms of action of a host of biologically important eukaryotic transcription factors. In addition, the scientific community has acquired a large structural database of both prokaryotic and eukaryotic DNA-binding proteins bound to their cognate sequences. Analyses of these complexes suggest that proteins recognize specific DNA sequences via induced fit (69). Conformational changes occur at the protein and/or DNA level, yielding an extensive protein-DNA interface composed of complementary interaction surfaces without bad steric contacts. Thermodynamic measurements have documented that protein-DNA recognition characteristically causes a dramatic negative heat capacity change. This thermodynamic signature reflects the large favorable entropy changes that accompany complex formation, because of reorganization of water and electrolytes and changes in conformational and vibrational flexibility of the two macromolecules.

The immediate challenge facing structural biologists and biophysicists studying protein-DNA complexes will be to go beyond the thermodynamic considerations enumerated above and develop a detailed understanding for the stereochemical bases of the large negative heat capacity change observed during molecular recognition. At a minimum, the induced-fit model should encourage structural and thermodynamic studies of biologically important transcription factor–DNA complexes. Fortunately, we can expect to learn a great deal from protein-protein molecular recognition events, which manifest the same structural and thermodynamic properties (i.e. chemical and shape complementarity, and large negative heat capacity change; reviewed in 56, 59). In the longer term, we will have to use tools from chemistry and physics for rigorous studies of dynamic aspects of molecular recognition. We must understand sequence-dependent DNA deformability in quantitative terms. Full knowledge of the kinetics of protein-DNA association and dissociation will also be essential if we hope to develop a detailed mechanistic understanding of regulated gene expression in eukaryotes. Finally, we must also try to come to grips with the effects of macromolecular crowding in the nucleus (reviewed in 36).

### ACKNOWLEDGMENTS

We are grateful to the many structural biologists who provided us with reprints and access to unpublished material.

*Literature Cited*


1. Andre B. 1990. The UGA3 gene regulating the GABA catabolic pathway in Saccharomyces cerevisiae codes for a putative zinc finger protein acting on RNA levels. *Mol. Gen. Genet.* 220:269–76
2. Anthony-Cahill SJ, Benfield PA, Robert F, Wasserman ZR, Brenner SL, et al. 1992. Molecular characterization of helix-loop-helix peptides. *Science* 255:979–83
3. Assa-Munt N, Mortshire-Smith RJ, Aurora R, Herr W, Wright PE. 1993. The solution structure of the Oct-1 POU-specific domain reveals striking similarity to the bacteriophage λ repressor DNA-binding domain. *Cell* 73:193–205
4. Baeuerle PA, Henkel T. 1994. Function and activation of NF-κB in the immune system. *Annu. Rev. Immunol.* 12:141–79
5. Baldwin RL. 1986. Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. USA* 83:8069–72
6. Barberis A, Pearlberg J, Simkovich N, Farrell S, Reinagle P, et al. 1995. Contact with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* 81:359–68
7. Barlow PN, Luisi B, Milner A, Elliot M, Everett R. 1994. Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc finger. *J. Mol. Biol.* 237:201–11
8. Berg JM. 1991. Zinc finger domains: hypotheses and current knowledge. *Annu. Rev. Biophys. Biophys. Chem.* 19:405–21
9. Billeter M, Qian YQ, Otting G, Muller M, Gehring W, et al. 1993. Determination of the nuclear magnetic solution structure of an Antennapedia homeodomain-DNA complex. *J. Mol. Biol.* 234:1084–97
10. Burley SK, Roeder RG. 1996. Biochemistry and structural biology of transcription factor IID. *Annu. Rev. Biochem.* 65:769–99
11. Canne LE, Ferré-D'Amaré AR, Burley SK, Kent SBH. 1995. Total chemical synthesis of a unique transcription factor–related protein: cMyc-Max. *J. Am. Chem. Soc.* 117:2998–3007
12. Chen X, Ramakrishnan B, Sambhorao T, Sundaralingam M. 1994. Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nature Struct. Biol.* 1:169–75
13. Chi T, Lieberman P, Ellwood K, Carey M. 1995. A general mechanism for transcriptional synergy by eukaryotic activators. *Nature* 377:254–57
14. Chiang S-Y, Welch J, Rauscher F, Beerman T. 1994. Effects of minor groove binding drugs on the interaction of TATA box binding protein and TFIIA with DNA. *Biochemistry* 33:7033–40
15. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265:346–55
16. Choy B, Green MR. 1993. Eukaryotic activators function during multiple steps of preinitiation complex assembly. *Nature* 366:531–36
17. Clark KL, Halay ED, Lai E, Burley SK. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364:412–20
18. Cohen C, Parry DA. 1994. Alpha-helical coiled coils: more facts and better predictions. *Science* 263:488–89
19. Coll M, Aymami J, Marel G, Boom J, Wang AH-J. 1989. Molecular structure of the netropsin-d(CGCGATATCGCG) complex: DNA conformation in an alternating AT segment. *Biochemistry* 28:310–20
20. Crick FHC. 1953. The packing of α-helices: simple coiled-coils. *Acta Crystallogr.* 6:689–97
21. Damberger FF, Pelton JG, Liu C, Cho H, Harrison CJ, et al. 1995. Refined solution structure and dynamics of the DNA-binding domain of the heat shock factor from Kluyveromyces lactis. *J. Mol. Biol.* 254:704–19
22. Dekker N, Cox M, Boelens R, Verrijzer CP, van der Vleit PC, et al. 1993. Solution structure of the POU-specific domain of Oct-1. *Nature* 362:852–55
23. Donaldson LW, Petersen JM, Graves BJ, McIntosh LP. 1996. Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J.* 15:125–34
24. Drew H, Travers A. 1985. DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* 186:773–90
25. Ellenberger T. 1994. Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. *Curr. Opin. Struct. Biol.* 4:12–21
26. Ellenberger T, Fass D, Arnaud M,

Harrison SC. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* 8:970–80
27. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted $\alpha$-helices: crystal structure of the protein-DNA complex. *Cell* 71:1223–37
28. Engelke DR, Ng S-Y, Shastry BS, Roeder RG. 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* 19:717–28
29. Fairall L, Schwabe JW, Chapman L, Finch JT, Rhodes D. 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366:483–87
30. Ferré-D'Amaré AR, Burley SK. 1995. DNA recognition by helix-loop-helix proteins. *Nucleic Acids Mol. Biol.* 9:285–98
31. Ferré-D'Amaré AR, Pognonec P, Roeder RG, Burley SK. 1994. Structure and function of the b/HLH/Z domain of USF. *EMBO J.* 13:180–89
32. Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley SK. 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363:38–45
33. Finkelstein AV, Janin J. 1989. The price of lost freedom: entropy of bimolecular complex formation. *Prot. Eng.* 3:1–3
34. Fisher DE, Parent LA, Sharp PA. 1993. High affinity DNA-binding myc analogs: recognition by an $\alpha$-helix. *Cell* 72:467–76
35. Freedman LP, Luisi BF. 1993. On the mechanism of DNA binding by nuclear hormone receptor: a structural and functional perspective. *J. Cell. Biochem.* 51:140–50
36. Garner MM, Burg MB. 1994. Macromolecular crowding and confinement in cells exposed to hypertonicity. *Am. J. Physiol. Cell Physiol.* 266:877–92
37. Ge H, Roeder RG. 1994. Purification, cloning and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* 78:513–23
38. Geiger JH, Hahn S, Lee S, Sigler PB. 1996. The crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* 272:830–36
39. Ghosh G, Duyne GV, Ghosh S, Sigler PB. 1995. Structure of the NF-$\kappa$B p50 homodimer bound to a kB site. *Nature* 373:303–10
40. Gill G. 1994. Taking the initiative. *Curr. Biol.* 4:374–76
41. Glover JNM, Harrison SC. 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* 373:257–61
42. Gruss P, Walther C. 1992. Pax in development. *Cell* 69:719–22
43. Ha J-H, Spolar RS, Record TMJ. 1989. Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J. Mol. Biol.* 209:801–16
44. Hahn S. 1992. The Yin and Yang of mammalian transcription. *Curr. Biol.* 2:152–54
45. Ham J, Dostatni N, Gauthier JM, Yaniv M. 1991. The papillomavirus E2 protein: a factor with many talents. *Trends Biochem. Sci.* 16:440–44
46. Haran TE, Joachimiak A, Sigler PB. 1992. The DNA target of the trp repressor. *EMBO J.* 11:3021–30
47. Hard T, Kellenbach E, Boelens R, Maler BA, Dahlman K, et al. 1990. Solution structure of the glucocorticoid receptor DNA-binding domain. *Science* 249:157–60
48. Harris C, Hollstein M. 1993. Clinical implications of the p53 tumor-suppressor gene. *N. Engl. J. Med.* 329:1318–27
49. Harrison CJ, Bohm AA, Nelson HCM. 1994. Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science* 263:224–27
50. Hegde RS, Grossman SR, Laimins LA, Sigler PB. 1992. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* 359:505–12
51. Helmann J, Chamberlin M. 1988. Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* 57:839–72
52. Hirsch JA, Aggarwal AK. 1995. Structure of the even-skipped homeodomain complexed with AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J.* 14:6280–91
53. Hoopes B, LeBlanc J, Hawley D. 1992. Kinetic analysis of yeast TFIID-TATA box complex formation suggests a multistep pathway. *J. Biol. Chem.* 267:11539–46
54. Hori R, Carey M. 1994. The role of activators in assembly of RNA polymerase II transcription complexes. *Curr. Opin. Genet. Dev.* 4:236–44
55. Houbaviy HB, Usheva AA, Shenk T, Burley SK. 1996. Co-crystal structure

of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. USA.* 93:13577–82
56. Janin J. 1995. Elusive affinities. *Proteins* 21:30–39
57. Jencks WP. 1981. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. USA* 78:4046–50
58. Joachimiak A, Haran TE, Sigler PB. 1994. Mutagenesis supports water mediated recognition in the trp repressor-operator system. *EMBO J.* 13:367–72
59. Jones S, Thornton JM. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* 93:13–20
60. Kim JL, Burley SK. 1994. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct. Biol.* 1:638– 53
61. Kim JL, Nikolov DB, Burley SK. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365:520–27
62. Kim Y, Geiger JH, Hahn S, Sigler PB. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365:512–20
63. Kissinger CR, Liu B, Martin-Blanco E, Kornberg TB, Pabo CO. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* 63:579–90
64. Klemm JD, Rould MA, Aurora R, Herr W, Pabo CO. 1994. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding domains. *Cell* 77:21–32
65. Kodandapani R, Pio F, Ni C-Z, Piccailli G, Klemsz M, et al. 1996. A new pattern for helix-turn-helix recognition revealed by the PU.1 Ets-domain-DNA complex. *Nature* 380:456–59
66. Koleske A, Young R. 1995. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* 20:113–16
67. König P, Richmond TJ. 1993. The X-ray structure of the GCN4-bZip bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* 233:139–54
68. Kopka M, Yoon C, Goodsell D, Pjura P, Dickerson R. 1985. The binding of an antitumor drug to DNA. Netropsin and CGCGAATTBrCGCG. *J. Mol. Biol.* 183:553–63
69. Koshland DE Jr. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA* 44:98–114
70. Kraulis PJ. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24:946–50
71. Ladbury JE, Wright JG, Sturtevant JM, Sigler PB. 1994. A thermodynamic study of the trp repressor-operator interaction. *J. Mol. Biol.* 238:669–81
72. Lai E, Clark KL, Burley SK, Darnell JE. 1993. Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: A family of transcription factors of diverse biological function. *Proc. Natl. Acad. Sci. USA* 90:10421–23
73. Landschultz WH, Johnson PF, McKnight SL. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA-binding proteins. 240:1759–64
74. Lawson CL, Carey J. 1993. Tandem binding in crystals of a trp repressor/operator half-site complex. *Nature* 366:178–82
75. LeBowitz JH, Clerc RG, Brenowitz M, Sharp PA. 1989. The Oct-2 protein binds cooperatively to adjacent octamer sites. *Genes Dev.* 3:1625–38
76. Lee MS, Gippert GP, Soman KV, Case DA, Wright PE. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245:635–37
77. Li T, Stark MR, Johnson AD, Wolberger C. 1995. Crystal structure of the Mata1/Mata2 homeodomain heterodimer bound to DNA. *Science* 270: 262–69
78. Li Y, Flanagan PM, Tschochner H, Kornberg RD. 1994. RNA polymerase II initiation factor interactions and transcription start site selection. *Science* 263:805–7
79. Liang H, Mao X, Olenjniczak ET, Nettesheim DG, Yu L, et al. 1994. Solution structure of the ets domain of Fli-1 when bound to DNA. *Nature Struct. Biol.* 1:871–75
80. Livingstone J, Spolar R, Record MJ. 1991. Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry* 30:4237–44
81. Love JJ, Li X, Case DA, Giese K, Grosschedl R, et al. 1995. Structural basis by DNA bending by the architectural transcription factor LEF-1. *Nature* 376:791–95
82. Luisi BF, Xu WX, Otwinowski Z,

Feedman LP, Yamamoto KR, et al. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352:497–505
83. Ma PCM, Rould MA, Weintraub H, Pabo CO. 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77:451–59
84. Marmorstein R, Carey M, Ptashne M, Harrison SC. 1992. DNA recognition by Gal4: structure of a protein-DNA complex. *Nature* 356:408–14
85. Marmorstein R, Harrison SC. 1994. Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a $Zn^2Cys^6$ binuclear cluster. *Genes Dev.* 8:2504–12
86. Matsui T, Segall J, Weil P, Roeder R. 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem.* 255:11992–96
87. Matthews BW. 1988. Protein-DNA interaction. No code for recognition. *Nature* 335:294–95
88. Meisterernst M, Horikoshi M, Roeder RG. 1990. Recombinant yeast TFIID, a general transcription factor, mediates activation by the gene-specific factor USF in a chromatin assembly assay. *Proc. Natl. Acad. Sci. USA* 87:9153–57
89. Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus ooctyes. *EMBO J.* 4:1609–14
90. Muller CW, Rey FA, Sodeoka M, Verdine GL, Harrison SC. 1995. Structure of the NF-$\kappa$B p50 homodimer bound to DNA. *Nature* 373:311–17
91. Murre C, McCaw PS, Baltimore D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56:777–83
92. Nekludova L, Pabo C. 1994. Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes. *Proc. Natl. Acad. Sci. USA* 91:6948–52
93. Nikolov DB, Burley SK. 1994. 2.1 Å Resolution refined structure of a TATA box-binding protein (TBP). *Nature Struct. Biol.* 1:621–37
94. Nikolov DB, Chen H, Halay E, Usheva A, Hisatake K, et al. 1995. Crystal structure of a TFIIB-TBP-TATA element ternary complex. *Nature* 377:119–28
95. Nikolov DB, Chen H, Halay ED, Hoffmann A, Roeder RG, et al. 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. USA* 93:4956–61
96. Norman C, Runswick M, Pollock R, Treisman R. 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55:989–1003
97. O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254:539–44
98. Ogata K, Hojo H, Aimoto S, Nakai T, Nakamura H, et al. 1992. Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl. Acad. Sci. USA* 89:6428–32
99. Ogata K, Morikawa S, Nakamura H, Serikawa A, Inoue T, et al. 1994. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* 79:639–48
100. Omichinski JG, Clore GM, Schaad O, Felsenfeld G, Trainor C, et al. 1993. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* 261:438–46
101. Orkin SH. 1992. GATA-binding transcription factors in hematopoietic cells. *Blood* 80:575–81
102. Otting G, Qian YQ, Billeter M, Muller M, Affolter M, et al. 1990. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy. *EMBO J.* 9:3085–92
103. Otwinowski Z, Schevitz RW, Zhang R-G, Lawson CL, Joachimiak A, et al. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335:321–29
104. Pabo CO, Sauer RT. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61:1053–95
105. Paolella DN, Palmer CR, Schepartz A. 1994. DNA targets for certain BZIP proteins distinguished by an intrinsic bend. *Science* 264:1130–33
106. Parkhurst K, Brenowitz M, Parkhurst L. 1996. Simultaneous binding and bending of promoter DNA by TBP: real-time kinetic measurements. *Biochemistry* 35:7459–65
107. Parvin J, McCormick R, Sharp P, Fisher D. 1995. Pre-bending of a promoter se-

quence enhances affinity for the TATA-binding factor. *Nature* 273:724–27
108. Parvin J, Sharp P. 1993. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* 73:533–40
109. Deleted in proof.
110. Pavletich NP, Pabo CO. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809–17
111. Pavletich NP, Pabo CO. 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* 261:1701–7
112. Pellegrini L, Tan S, Richmond TJ. 1995. Structure of serum response factor core bound to DNA. *Nature* 376:490–98
113. Pinto I, Wu W-H, Na JG, Hampsey M. 1994. Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. *J. Biol. Chem.* 269:30569–73
114. Prioleau M-N, Huet J, Sentenac A, Mechali M. 1994. Competition between chromatin and transcription complex assembly regulates gene expression during early development. *Cell* 77:439–49
115. Qian X, Jeon CJ, Yoon HS, Agarwal K, Weiss MA. 1993. Structure of a new nucleic-acid-binding motif in eukaryotic transcription elongation factor TFIIA. *Nature* 365:277–79
116. Qian YQ, Otting G, Billeter M, Müller M, Gehring W, et al. 1993. Nuclear magnetic resonance spectroscopy of a DNA complex with the uniformly 13C labeled antennapedia homeodomain and structure determination of the DNA-bound homeodomain. *J. Mol. Biol.* 234:1070–83
117. Qian YQ, Otting G, Wüthrich K. 1993. NMR detection of hydration water in the intermolecular interface of a protein-DNA complex. *J. Am. Chem. Soc.* 115:1189–90
118. Ramakrishnan V, Finch J, Graziano V, Sweet R. 1993. Crystal structure of the globular domain of histone H5 and its implications for nucleosome binding. *Nature* 362:219–23
119. Rastinejad F, Perlmann T, Evans RM, Sigler PB. 1995. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* 375:203–11
120. Record TM, Ha J-H, Fisher MA. 1991. Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Meth. Enzymol.* 208:291–343
121. Rigaud G, Roux J, Pictet R, Grange T. 1991. In vivo foot-printing of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific transcription factor. *Cell* 67:977–86
122. Roeder RG. 1996. The role of general initiation factors by RNA polymerase II. *Trends Biochem. Sci.* 21:327–35
123. Rosenfeld MG. 1991. POU-domain transcription factors: pou-er-ful developmental regulators. *Genes Dev.* 5:897–907
124. Satchwell S, Drew H, Travers A. 1986. Sequence periodicities in chicken nucleosomal core DNA. *J. Mol. Biol.* 191:659–79
125. Saudek V, Pastore A, Castiglion-Morelli MA, Frank R, Gausepohl H, et al. 1991. The solution structure of a leucine-zipper motif peptide. *Protein Eng.* 4:519–29
126. Sauer F, Hansen S, Tjian R. 1995. Multiple TAFIIs directing synergistic activation of transcription. *Science* 270:1783–88
127. Sauer F, Hansen S, Tjian R. 1995. DNA template and activator-coactivator requirements for transcriptional synergism by Drosophila bicoid. *Science* 270:1825–28
128. Sawadogo M, Roeder RG. 1985. Interaction of a gene-specific transcription factor with the Adenovirus major late promoter upstream of the TATA box region. *Cell* 43:165–75
129. Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science* 253:1001–7
130. Schwabe JWR, Chapman L, Finch JT, Rhodes D. 1993. The crystal structure of the oestrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* 75:567–78
131. Schwabe JWR, Chapman L, Finch JT, Rhodes D. 1993. DNA recognition by the oestrogen receptor: from solution to the crystal. *Structure* 1:187–204
132. Schwabe JWR, Chapman L, Rhodes D. 1995. The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure* 3:201–13
133. Schwabe JWR, Klug A. 1994. Zinc mining for protein domains. *Nature Struct. Biol.* 1:345–50
133a. Schwabe JWR, Neuhans D, Rhodes D.

1990. Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* 348:458–461
134. Seto E, Shi Y, Shenk T. 1991. YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* 354:241–45
135. Shakked Z, Guzikevish-Guerstein G, Frolow F, Rabinovich D, Joachimiak A, et al. 1994. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* 368:469–73
136. Shi Y, Berg J. 1995. Specific DNA-RNA hybrid binding by zinc-finger proteins. *Science* 268:282–84
137. Shi Y, Berg JM. 1995. DNA unwinding induced by zinc finger protein binding. *Biochemistry* 35:3845–48
138. Spolar RS, Record TM. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263:777–84
139. Staacke D, Walter B, Kisters-Woike K, Wilcken-Bergmann BV, Muller-Hill B. 1990. How trp repressor binds its operator. *EMBO J.* 9:1963–67
140. Steitz TA. 1990. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* 23:105–80
141. Sturtevant JM. 1977. Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci. USA* 74:2236–40
142. Sun D, Hurley L. 1995. TBP unwinding of the TATA box induces a specific downstream unwinding site that is targeted by pluramycin. *Chem. Biol.* 2:457–69
143. Tan S, Hunziker Y, Sargent DF, Richmond TJ. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* 381:127–34
144. Tjian R, Maniatis T. 1994. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77:5–8
145. Usheva A, Shenk T. 1994. TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* 76:1115–21
146. Usheva A, Shenk T. 1996. TFIIB and the large subunit of RNA polymerase II bind directly to YY1 to mediate transcription initiation. *Proc. Natl. Acad. Sci. USA.* 93:13571–76
147. Vinson CR, Sigler PB, Mcknight SL. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* 246:911–16
148. Weiss MA, Ellenberger T, Wobbe RC, Lee JP, Harrison SC, et al. 1990. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* 347:575–78
149. Werner M, Huth J, Gronenborn A, Clore M. 1995. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional structure of the human SRY-DNA complex. *Cell* 81:705–14
150. Werner MH, Clore GM, Fisher CL, Fisher RJ, Trihn L, et al. 1996. The solution structure of the human Ets1-DNA complex reveals a novel mode of DNA binding and true side chain intercalation. *Cell* 83:761–71
151. Wilson DS, Guenther B, Desplan C, Kuriyan J. 1995. Crystal structure of a paired (PAX) class cooperative homeodomain dimer on DNA. *Cell* 82:709–19
152. Wolberger C, Vershon AK, Liu B, Johnson AD, Pabo C. 1991. Crystal structure of a MAT $\alpha$2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* 67:517–28
153. Workman JL, Roeder RG. 1987. Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. *Cell* 51:613–22
154. Xu W, Rould MA, Jun S, Desplan C, Pabo CO. 1995. Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* 80:639–50
155. Zawel L, Kumar K, Reinberg D. 1995. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* 9:1479–90
156. Zhu WL, Zeng QD, Colangelo CM, Lewis LM, Summers MF, et al. 1995. The N-terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nature Struct. Biol.* 3:122–24