Cell, Vol. 88, 733–736, March 21, 1997, Copyright ©1997 by Cell Press

Architectural Transcription Factors: Proteins That Remodel DNA

Minireview

Milton H. Werner* and Stephen K. Burley*†
*The Rockefeller University
†Howard Hughes Medical Institute
1230 York Avenue
New York, New York 10021

During the past four decades, an enormous number of transcriptionally active proteins have been identified. Although some viral RNA polymerases are single polypeptide chains that are intrinsically capable of RNA synthesis, the cellular enzymes are much more complicated. These multisubunit, DNA-dependent RNA polymerases require a host of accessory factors to ensure correct positioning of the enzyme at the transcription start site and control of the number of copies of RNA manufactured per time unit in response to developmental and environmental signals. Accessory molecules include sequence-specific DNA-binding factors and other proteins that together assemble on DNA control regions, or promoters, where they support RNA synthesis. In favorable cases, in vitro transcription systems using highly purified and/or recombinant components have been established, permitting detailed mechanistic studies of particular steps in these enormously complicated processes. Moreover, structural biologists have determined high-resolution three-dimensional structures of promoter-recognition factors (E. coli σ factor, TATA box–binding protein, Yin and Yang 1), a eukaryotic general transcription initiation factor (transcription factor IIB), transcription elongation factors (GreA, transcription factor IIS), and transcriptional activators and coactivators (E. coli cAMP receptor protein, transcription factor IIA, TATA box–binding protein associated factors, plus more than twenty examples of eukaryotic transcriptional activators bound to their DNA targets).

Each X-ray or NMR structure contains a tremendous wealth of atomic detail, which is of immediate relevance to biologists focusing on the mechanistic role of that particular transcription factor. An intriguing subset of our growing structural database tells us something more general about the three-dimensional organization of macromolecular machines responsible for transcription initiation. A handful of recently determined structures reveal proteins (architectural transcription factors) bound to either the major or the minor groove faces of DNA molecules in which kinks have been induced, dramatically altering the trajectory of the double helix (reviewed in Patikoglou and Burley, 1997). Together, these striking examples suggest that some transcription complexes are structurally much more elaborate than a simple linear double helix decorated with RNA polymerase and transcriptional regulatory proteins (reviewed earlier in Tjian and Maniatis, 1994).

Examples from Prokaryotes

Studies of transcriptional activators in prokaryotes conclusively established that architectural transcription factors regulate gene expression. The best characterized example is the cAMP receptor protein (CAP), which binds as a homodimer to a 22 bp pseudopalindromic DNA sequence found in many bacterial promoters immediately upstream of the -10 and -35 regions recognized by σ factor. In a landmark X-ray crystallographic study, Steitz and coworkers (Schultz et al., 1991) provided the first three-dimensional picture of a protein distorting the structure of DNA (Figure 1a). Each CAP monomer presents an α helix to the major groove of its cognate half site where it imparts an $\sim$45° kink, thereby introducing an overall bend of about 90° within two helical turns of DNA. Visualization of a protein-induced DNA bend provided an elegant mechanical explanation for solution biochemical findings, which suggested that CAP-bound DNA was not linear (Zinkel and Crothers, 1990).

Binding of the CAP dimer results in a smooth bend in the DNA with a concomitant compression of the minor groove in the vicinity of the bend center (Schultz et al., 1991). The double helix wraps around the sides of the protein dimer, creating a stable protein–DNA complex that presents at least one CAP subunit to the RNA polymerase (RNAP) holoenzyme bound to the core promoter. Although we do not yet have a structure for the CAP-RNAP-DNA triple complex, site-directed mutagenesis has yielded a fairly detailed picture of the molecular arrangement within this activated transcription complex. In vivo analysis of CAP mutants that do not stimulate transcription, but are capable of both DNA binding and bending, identified a region of CAP responsible for activation that is far removed from the DNA-binding surface. Construction of CAP heterodimers in which one subunit was functional for DNA-binding but not activation and the other subunit functional in activation but not DNA-binding established that only one of the CAP monomers participated in transcriptional activation—viz., the promoter-proximal subunit of the CAP dimer (reviewed in Ebright, 1995). The protein-induced DNA bend coupled with the observation that CAP activation involves the CAP subunit located closest to the binding site for RNAP suggested that the so-called activating region of CAP stimulates transcription by direct interaction with RNAP. Mutational analysis of RNAP identified a region of its α subunit that is functionally important for CAP-dependent activation of transcription (Zou et al., 1992). This region of RNAPα has also been implicated in the activating effects of a number of other bacterial transcription factors, such as FNR, OmpR, and AraC.

Although most experimental work has been restricted to a subset of CAP-dependent promoters, it is believed that activation of transcription by CAP results from direct protein–protein interactions with RNAP creating the molecular arrangement modeled in Figure 1a (reviewed in Ebright, 1995). Thus, DNA deformation (a requirement for CAP binding) facilitates assembly of a higher-order, stereo-specific transcription complex that supports increased rates of transcription initiation.

Integration host factor (IHF) is a ubiquitous architectural regulator of bacterial physiology. First discovered as a host factor that supports bacteriophage λ integration, IHF participates in various processes dependent

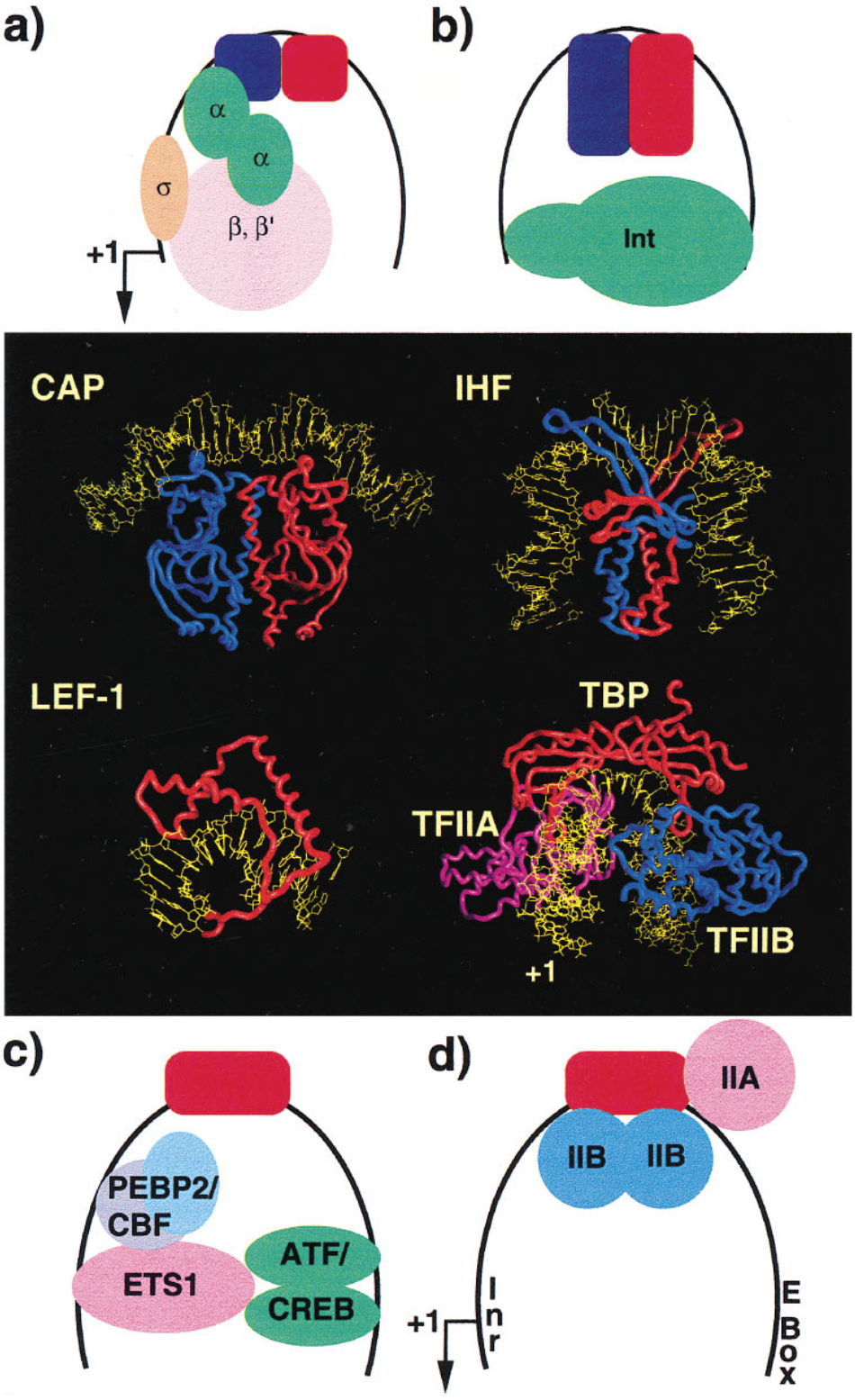

Figure 1. A Composite View of Architectural Transcription Factors in Action

on assembly of higher-order structures of protein and DNA, including transcriptional regulation (reviewed in Friedman, 1988) and recombination (reviewed in Nash, 1996). IHF consists of two structurally similar subunits (Figure 1b), which bind to an unusually long 30 bp binding site. The subunits of IHF are closely related to another bacterial protein known as HU; however, unlike IHF, HU cannot recognize specific DNA sequences.

IHF and HU form a structurally related family of proteins capable of producing a remarkable bend in DNA, estimated from earlier solution studies to exceed 140° (Thompson and Landy, 1988) and recently revealed by X-ray crystallography of an IHF-DNA complex (Figure 1b) to be a staggering 160°–180° (Rice et al., 1996). This unprecedented distortion is achieved via a unique mechanism of DNA binding involving a bipartite structure of arm-like β hairpin structures and C-terminal α helices, which make only minor groove contacts with DNA bases and the sugar-phosphate backbone. One of the most intriguing features of the IHF homodimer structure is the presence of two intercalating proline residues projecting from the tip of each β hairpin arm, which introduce large kinks at symmetrically displaced ApA steps each located approximately a half turn from of the center of the recognition site. This unusual mode of DNA distortion is shared by a number of structurally unrelated protein families, each of which insert nonpolar residues from the minor groove face into base steps yielding dramatic kinks in the trajectory of the double helix (see below).

The structure of the apo-form of HU (Tanaka et al., 1984) is very similar to that of IHF observed in the IHF complex, suggesting a general DNA-bending role for this structural motif. Moreover, in vitro HU can substitute for IHF during site-specific recombination (reviewed in Nash, 1996), reinforcing the concept that this structural motif mediates assembly of higher-order nucleoprotein complexes. A hypothetical intasome assembly of IHF with bacteriophage λ integrase (Int) is illustrated in Figure 1b (adapted from Nash, 1996).

Examples from Eukaryotes

Three architectural transcription factors from eukaryotes have been examined at high resolution by structural biologists, including the TATA box-binding protein (TBP), SRY, and LEF-1. Structures of these three proteins bound to their respective DNA targets revealed common strategies for DNA deformation, analogous to the minor groove intercalation events seen in the IHF-DNA cocrystal structure. Furthermore, these proteins create stereo-specific protein–nucleic acid complexes that are in turn recognized by other transcription factors.

SRY (Werner et al., 1995) and LEF-1 (Love et al., 1995) are structurally similar proteins that both use a nonaromatic, hydrophobic amino acid to intercalate into a base step from the minor groove face of the double helix, creating a dramatically bent DNA (∼130° in the case of LEF-1) with a widened minor groove (Figure 1c). They are high mobility group (HMG) proteins that possess a DNA-binding motif first described in the nonhistone chromosomal proteins HMG1 and HMG2 (reviewed in Werner et al., 1995), which are abundant nuclear proteins that exhibit nonspecific DNA binding and are often referred to as eukaryotic homologs of HU (see above).

LEF-1 is a pre-B and T lymphocyte–specific DNA-binding protein capable of nucleating formation of a large macromolecular complex, or enhanceosome, that stimulates transcription of various lymphocyte-specific genes (reviewed in Giese et al., 1995). Assembly of this enhanceosome depends on the particular arrangement of factor-binding sites and proteins, which is established through protein–protein and protein–DNA interactions. The best characterized example of a DNA segment capable of supporting assembly of a multiprotein–DNA complex is provided by the minimal T cell receptor (TCR) α gene enhancer, which functions specifically in T lymphocytes. The enhancer is constructed from at least four distinct factor binding sites, with the LEF-1 recognition site at its center (Figure 1c). LEF-1-induced DNA bending enables interaction of at least two additional enhancer-bound proteins. Without the bend, ETS1, PEBP2/CBF and ATF/CREB would be far from one another in space. Assembly of a functional TCR α gene enhancer complex, therefore, requires DNA remodeling by LEF-1 creating the foundation on which a specific multiprotein–DNA complex is constructed (reviewed in Love et al., 1995).

TBP is a highly conserved, central transcription factor required for initiation of RNA synthesis by all three eukaryotic RNA polymerases (reviewed in Tansey and Herr, 1997). RNA polymerase II (pol II) transcription begins with TBP recognizing a conserved AT-rich promoter element known as the TATA box. TBP-DNA complexes derived from plant, yeast, and human reveal two pairs of phenylalanine residues intercalating into base steps creating two ∼45° kinks at either end of the TATA element (reviewed in Patikoglou and Burley, 1997). These intercalation events also mediate abrupt transitions to and from canonical B form DNA and an underwound, smoothly bent DNA structure in which the widened minor groove is accessible to the antiparallel β sheet DNA-binding surface of TBP (Figure 1d). The path of the DNA axis resembles an incomplete right-handed supercoil, which compensates for partial unwinding of the right-handed double helix and appears to prevent helical strain of the promoter. The resulting nucleoprotein complex serves as an extremely stable platform on which the pol II transcription machinery is assembled.

Further X-ray studies have yielded structures of two triple complexes, transcription factor (TF) IIB-TBP-DNA (Nikolov et al., 1995) and TFIIA-TBP-DNA (Geiger et al., 1996; Tan et al., 1996), which have been combined in Figure 1d to give a model for the TFIIA, TFIIB, and TBP bound to the adenovirus major late promoter (AdMLP). Both triple complex structures demonstrate the same DNA deformation seen in the TBP-DNA cocrystal structures, suggesting that TFIIB and TFIIA recognize the preformed TBP-DNA complex via specific protein–protein interactions with TBP and contacts with the phosphodeoxyribose backbone of the distorted double helix. In addition to providing the platform for pol II preinitiation complex assembly, the TBP-DNA complex may serve an architectural role analogous to IHF (Figure 1b) or LEF-1 (Figure 1c) by bringing remote binding sites for transcriptional activators, such as the E box, closer to the site of action of pol II (Figure 1d). There is evidence that upstream stimulatory factor (USF) functions through both the Initiator element (Inr) and the E box of the AdMLP, possibly via DNA looping mediated by USF tetramerization (reviewed in Ferre-D'Amare et al., 1994).

In contrast to CAP, IHF, SRY/LEF-1, and TBP, there is at least one architectural transcription factor that may function by remodeling a preexisting DNA deformation. HMG I(Y) is a nonhistone chromosomal protein, which despite its name does not appear to contain an HMG1/

2-type DNA-binding domain (reviewed in Falvo et al., 1995). Within the interferon β (IFNβ) gene enhanceosome, HMG I(Y) is thought to facilitate assembly of a stereo-specific protein–nucleic acid complex by simultaneously occupying the DNA element recognized by NF-κB. Electrophoretic mobility shift analyses suggest that HMG I(Y) reduces the intrinsic curvature of this DNA segment, thereby stabilizing NF-κb binding and assembly of the IFNβ enhanceosome (Falvo et al., 1995).

Conclusion

Structural biologists have made considerable progress toward defining the physical chemistry of complexes between architectural transcription factors and remodeled DNA. The quest to crystallize even larger multiprotein–DNA assemblies continues apace, and we should soon see actual three-dimensional structures of the CAP-DNA complex interacting with a portion of the α subunit of RNAP (modeled in Figure 1a), a quaternary complex of the TCR α gene enhancer being recognized by LEF-1, ETS1, and ATF/CREB (modeled in Figure 1c), and TFIIA-, TFIIB-, and TBP-bound DNA (modeled in Figure 1d). The immediate challenge facing molecular biologists is to go beyond mere physical descriptions of these systems and characterize the kinetic and thermodynamic consequences of building transcriptionally active stereo-specific nucleoprotein complexes. Success in these difficult endeavors should yield quantitative insights into how gene expression is regulated, which are of critical importance to fundamental molecular biology as well as biomedicine.

Selected Reading

Ebright, R. (1995). Mol. Microbiol. *8*, 797–802.

Falvo, J.V., Thanos, D., and Maniatis, D. (1995) Cell *83*, 1101–1111.

Ferre-D'Amare, A.R., Pognonec, P., Roeder, R.G., and Burley, S.K. (1994). EMBO J. *13*, 180–189.

Friedman, D.I. (1988). Cell *55*, 545–554.

Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996). Science *272*, 830–838.

Giese, K., Kingsley, C., Kirshner, J.R., and Grosschedl, R. (1995). Genes Dev. *9*, 995–1008.

Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R., and Wright, P.E. (1995). Nature *376*, 791–795.

Nash, H.A. (1996). In Regulation of Gene Expression in E. coli, E.C.C. Lin and A. Simon Lynch, eds. (Austin, TX: R. G. Landes Company), 149–179.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Nature *377*, 119–128.

Patikoglou, G.A., and Burley, S.K. (1997). Annu. Rev. Biophys. Biomol. Struct. *26*, 287–323.

Rice, P.A., Yang, S.-W., Mizuuchi, K., and Nash, H.A. (1996). Cell *87*, 1295–1306.

Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Science *253*, 1001–1007.

Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996). Nature *381*, 127–134.

Tanaka, I., Appelt, K., Dijk, J., White, S.W., and Wilson, K.S. (1984). Nature *310*, 376–381.

Tansey, W.P., and Herr, W. (1997). Cell, this issue.

Thompson, J., and Landy, A. (1988). Nucleic Acids Res. *16*, 9687–9705.

Tjian, R., and Maniatis, T. (1994). Cell *77*, 5–8.

Werner, M.H., Huth, J.R., Gronenborn, A.M., and Clore, M.G. (1995). Cell *81*, 705–714.

Zinkel, S., and Crothers, D. (1990). Biopolymers *29*, 29–38.

Zou, C., Fujita, N., Igarashi, K., and Ishihama, A. (1992). Mol. Microbiol. *6*, 2599–2605.