

Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator

HRISTO B. HOUBAVIY*, ANNY USHEVA†, THOMAS SHENK†‡, AND STEPHEN K. BURLEY*§¶

*Laboratories of Molecular Biophysics and §Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399; and †Department of Molecular Biology and ‡Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544-1014

Contributed by Thomas Shenk, September 12, 1996

ABSTRACT Ying–Yang 1 protein (YY1) supports specific, unidirectional initiation of messenger RNA production by RNA polymerase II from two adjacent start sites in the adeno-associated virus P5 promoter, a process which is independent of the TATA box-binding protein (TBP). The 2.5-Å resolution YY1-initiator element cocrystal structure reveals four zinc fingers recognizing a YY1-binding consensus sequence. Upstream of the transcription start sites protein–DNA contacts involve both strands and downstream they are virtually restricted to the template strand, permitting access to the active center of RNA polymerase II and ensuring specificity and directionality. The observed pattern of protein–DNA contacts also explains YY1 binding to a preformed transcription bubble, and YY1 binding to a DNA/RNA hybrid analog of the P5 promoter region containing a nascent RNA transcript. A model is proposed for YY1-directed, TBP-independent transcription initiation.

Transcription by RNA polymerase II (pol II) is controlled by TATA (1, 2) and/or pyrimidine-rich initiator (3) elements that function independently or synergistically in class II nuclear gene promoters (reviewed in ref. 4). Accurate transcription initiation depends on assembling pol II and general transcription initiation factors on the promoter, yielding a preinitiation complex (PIC; reviewed in ref. 5). In the most general case, PIC assembly begins with the TATA box-binding protein (TBP) subunit of transcription factor IID (TFIID) recognizing the TATA element, followed by TFIIB recruitment, creating a TBP–TFIIB–DNA ternary complex that directs accretion of pol II and TFIIF, TFIIE, and TFIIH (reviewed in ref. 4). Once the PIC is assembled, strand separation occurs, and pol II becomes phosphorylated, initiates transcription, and is released from the promoter. During elongation *in vitro*, TFIID can remain bound to the promoter supporting rapid reinitiation of transcription (reviewed in ref. 6).

Ying–Yang 1 (YY1) is a human GLI–Kruppel-related protein capable of transcriptional activation or repression and sequence-specific initiator element binding (reviewed in ref. 7). *In vitro* studies demonstrated that YY1 recognition of the Initiator element of the adeno-associated virus (AAV) P5 promoter supports transcription initiation from two adjacent start sites (8) (initiator-dependent transcription is reviewed in ref. 9). Despite the presence of a TATA box in the P5 promoter, this process is independent of TBP (10). 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 (11). Specific, unidirectional, YY1-dependent initiation of transcription can proceed in the presence of nucleoside triphosphates if the P5 template is negatively supercoiled (10). 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 (12).

In this paper, we report the cocrystal structure of human YY1 recognizing the AAV P5 promoter initiator element. Our structure and the results of complementary biochemical studies permit detailed structural and functional analyses of the protein–DNA interactions, providing further insights into how YY1 may direct TBP-independent transcription initiation. Finally, a mechanistic model for YY1 function on the AAV P5 promoter is presented.

MATERIALS AND METHODS

Oligonucleotides and Gel Mobility Shift Assays. Oligonucleotides used for crystallization and gel mobility shift analyses were synthesized and purified as described (13). The gel mobility shift experiments were done according to ref. 14.

Protein Expression and Purification. Human YY1 residues 293 to 414 preceded by the dipeptide sequence ME (Δ YY1) was overexpressed in *Escherichia coli* (15). Denatured Δ YY1 was purified as described in ref. 16. Following lyophilization, the protein was reduced in 100 mM Hepes (pH 7.5), 6 M guanidinium-HCl, 100 mM dithiothreitol, and 100 μ M zinc acetate at \approx 5 mg/ml by incubating at 60°C for 30 min. Refolding was performed by diluting 1:100 with 50 mM Hepes (pH 7.5), 100 mM arginine-HCl, 5 mM MgCl₂, 5 mM dithiothreitol, and 100 μ M zinc acetate and incubating overnight at 10°C. Refolded Δ YY1 was further purified to homogeneity by heparin and cation exchange chromatography. Mass spectrometry documented that the Δ YY1 used for crystallization was neither modified nor proteolyzed (measured molecular mass, 14,156 \pm 2; predicted molecular mass, 14,156). Dynamic light scattering (molecular size detector, Protein Solutions, Charlottesville, VA) showed that the Δ YY1–DNA complex is both monomeric and monodisperse at 5 mg/ml (data not shown). Full-length human YY1 was purified as previously described (10).

Crystallization. Δ YY1 and crystallization DNA were mixed in a 1.56:1 molar ratio in 20 mM Hepes (pH 7.5), 100 mM NaCl, 100 mM ammonium acetate, 5 mM MgCl₂, 10 mM dithiothreitol, and 100 μ M zinc acetate and concentrated by ultrafiltration to a final Δ YY1 concentration of 0.9–1.2 mM (corresponding to 0.6–0.8 mM DNA). Cocrystals were grown at 4°C by vapor diffusion against a reservoir solution consisting of 100 mM Hepes, 100 mM NaCl, 5 mM MgCl₂, 25% PEG monomethyl ether (M_r , 5000). Orthorhombic crystals with one complex in the asymmetric unit (P2₁2₁2₁; a = 44.0 Å, b = 65.7 Å, c = 117.1 Å) and typical dimensions 0.6 \times 0.3 \times 0.05 mm³

Abbreviations: AAV, adeno-associated virus; PIC, preinitiation complex; TBP, TATA box-binding protein; Δ YY1, human YY1 residues 293–414; pol II, RNA polymerase II; rmsd, root-mean-square deviation. The atomic coordinates have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1UBD).

¶To whom reprint requests should be sent at the * address.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Summary of crystallographic analysis

	Native1	Native2	I-dU ⁵	I-dU ¹³	I-dU ²⁴	I-dU ²⁵
Resolution, Å	2.5	3.0	2.8	3.0	3.0	3.0
Reflections, observations/unique	64,023/12,347	99,488/7,275	89,523/8,962	65,181/7,208	66,596/7,287	62,038/7,237
Data coverage, %	93.6	93.4	97.0	95.4	80.8	88.4
R_{sym} , %	8.0	7.5	6.7	8.7	9.6	6.6
Mean fractional isomorphous difference, %			13.7	18.9	10.0	18.6
MIR analysis, 20.0–3.0 Å against Native2						
Phasing power			1.27	0.69	0.83	0.86
R_c			0.61	0.69	0.68	0.69
Mean overall figure of merit	0.62					
Refinement (against Native1)						
Resolution, Å	6.0–2.5					
R factor/free R factor, %	21.2/33.0					
Reflections, $ F > 2.0\sigma F $ working/test	9,726/1,121					
Total no. of atoms	1,809					
No. of water molecules	86					
rms bond length, Å	0.016					
rms bond angle, degrees	2.07					
rms B-factor bonded atoms, Å ²	2.29					

rms, Root mean square.

were used for data collection. Crystals of heavy atom derivatives were prepared in the same manner using DNA in which 5-iodouridine had been substituted for T.

Data Collection and Structure Determination. Diffraction measurements were carried out at -150°C . Lower resolution data (Native2 and all derivatives) were collected using $\text{CuK}\alpha$ radiation from a rotating anode x-ray source. Native data (2.5 Å resolution) were obtained on Beamline X12B of the National Synchrotron Light Source. Crystals were transiently “cryoprotected” in 15% PEG monomethyl ether (M_r , 350), 19% PEG monomethyl ether (M_r , 5000), 50 mM Hepes (pH 7.5), 100 mM NaCl, 5 mM MgCl_2 , 10 mM dithiothreitol, and 100 μM zinc acetate before freezing by immersion in liquid propane. Oscillation photographs were integrated, scaled, and merged using DENZO and SCALEPACK (Z. Otwinowski, personal communication). Iodine atoms were located by difference (isomorphous and anomalous) Patterson and Fourier methods. Refinement of the heavy atom model against both isomorphous and anomalous differences gave a figure of merit of 0.62 (17). The quality of the MIR electron density map was sufficient to permit building of most of the DNA model (Gua-4–Ade-17 and Thy-24–Cyt-37). The four zinc ions appeared as strong peaks, and there was clear density for fingers 1, 2, and 3. Combination of the MIR data and phases calculated from the partial DNA model yielded continuous density for the missing DNA portions, and most of the protein backbone, including finger 4. The phase-combined map was used to fit polyaniline models based on finger 1 of Zif268 (18) into the electron density corresponding to fingers 1, 2, and 4 of ΔYY1 , and a polyaniline model based on finger 2 of ΔYY1 was fit into the electron density corresponding to finger 2 of ΔYY1 . The partial model was subjected to rigid body and positional refinement with X-PLOR (19) against the Native2 data and successive iterations of model building, positional refinement and phase combination were done to trace regions with poor electron density. Amino acid side chains making DNA contacts were omitted until the final stages of this procedure. The model was then subjected to several rounds of positional refinement, simulated annealing and tightly restrained B-factor refinement with X-PLOR. The resolution limit was extended to 2.5 Å by refinement against data Native1. Local scaling of $|F_{\text{observed}}|$ against $|F_{\text{calculated}}|$ with LSCALE (M. Rould, personal communication) was done at the final stages of refinement to correct for anisotropic diffraction and absorption effects. During the final refinement stage, well-ordered water molecules were located using $(|F_{\text{observed}}| - |F_{\text{calculated}}|)$ difference

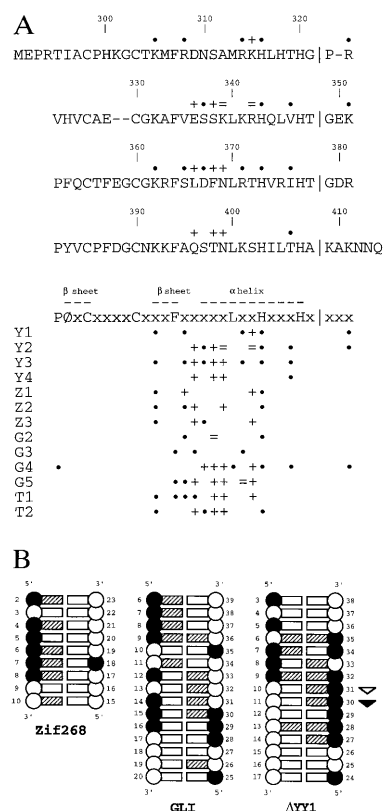


FIG. 1. Comparison of ΔYY1 with other zinc-finger structures. (A) Amino acid sequence of ΔYY1 . The zinc fingers are aligned one above the other, with the linker regions separated by vertical bars. Amino acids are numbered as in ref. 14. The ΔYY1 –DNA contacts (Y1–Y4, ●, backbone contact; +, base contact; =, touches both base and backbone) are also presented schematically with those seen in the structures of Zif268 (Z1–Z3) (18), GLI (G2–G5) (16), and Tramtrack (T1, T2) (21). Between the alignment panels, the zinc-finger consensus sequence (\emptyset , aromatic residue) and the conserved secondary structure elements are provided. Detailed comparisons of the ΔYY1 fingers with those of Zif268 and GLI reveal significant similarities [Y1 versus G3 root-mean-square deviation (rmsd) = 0.81 Å, Y2 versus Z3 rmsd = 0.61 Å, Y3 versus G3 rmsd = 0.73 Å, and Y4 versus G3 rmsd = 0.68 Å, which compare favorably with comparisons between Zif268 and GLI]. (B) Schematic comparison of the DNA contacts made by Zif268, GLI, and ΔYY1 . DNA bases and backbone are represented by rectangles and circles, respectively. Hatched rectangles and solid circles denote sites of protein contact.

Fourier syntheses. The current crystallographic model consists of amino acid residues 295–408, both DNA chains, four zinc ions, and 86 water molecules. The electron density for the polypeptide backbone is everywhere continuous at 1.0σ in a $(2|F_{\text{observed}}| - |F_{\text{calculated}}|)$ difference Fourier synthesis. PROCHECK (20) revealed no unfavorable backbone torsion angle combinations, and main-chain and side-chain structural parameters consistently better than those expected at 2.5-Å resolution (overall G-factor, 0.0). The DNA electron density is well defined throughout, except for the first and last two base pairs.

RESULTS AND DISCUSSION

Requirements for YY1-Dependent Transcription Initiation. Specific transcription initiation by pol II involves promoter recognition by one or more proteins that define both the start site and the direction of RNA synthesis. Neither TFIIB nor pol II are believed to be capable of significant sequence-specific interactions with DNA, which leaves YY1 as the only candidate for promoter recognition. *In vitro* studies demonstrated that YY1 recognizes the P5 Initiator element, and that the resulting YY1–DNA complex directs accretion of TFIIB and then pol II to the core promoter (11). Therefore, the YY1–

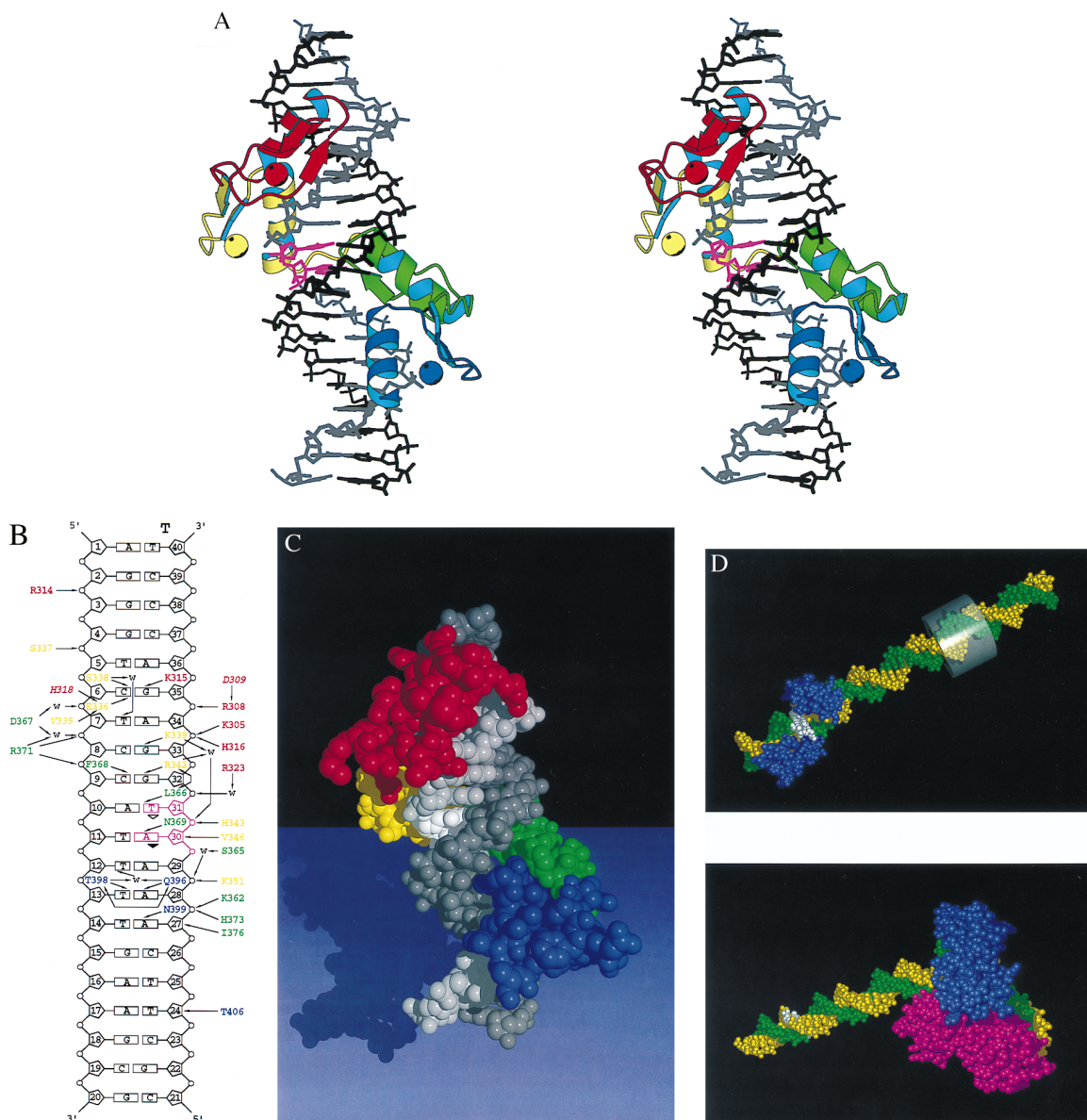


FIG. 2. Structure of the Δ YY1-initiator element complex. (A) Stereoview of the Δ YY1-DNA complex. The protein is shown as a ribbon representation, the DNA as a stick model, and the zinc ions as spheres. YY1 zinc fingers are colored red, yellow, green, and blue from N to C termini. The template strand of the DNA is colored grey, with the nucleotides corresponding to the two transcription start sites shown in purple, and the nontemplate strand is colored black. RNA synthesis proceeds downwards in this view. (B) Schematic representation of the Δ YY1-DNA interactions. The complete sequence of the crystallization oligonucleotide is shown, with the template strand labeled "T." Protein-DNA contacts are color coded as in A by their zinc finger of origin. Supporting interactions that stabilize side chains making DNA contacts are also shown in italic. The two transcription start sites are labeled with arrows denoting the direction of transcription (∇ and \blacktriangledown for dominant). Enthalpically favorable interactions include salt bridges ($<4 \text{ \AA}$), hydrogen bonds ($<3.5 \text{ \AA}$), van der Waals contacts ($<4 \text{ \AA}$), and a number of water-mediated bridges (denoted with "w"). (C) Space-filling representation of the Δ YY1-DNA complex. The protein and DNA are colored as in A, with the nucleotides corresponding to the two transcription start sites on the template strand shown in white. (D) Space-filling representations of the Δ YY1-P5 promoter complex (Upper) and the complex of TBP and the adenovirus major late promoter (AdMLP) (Lower). Δ YY1 and TBP are colored blue and the template and nontemplate strands of the DNA are colored green and yellow, respectively. The transcription start sites are denoted with white base pairs. A transparent cylinder surrounds the -24 to -31 region of the P5 promoter (the TATA element in the AdMLP).

initiator element complex must be capable of presenting the transcription start sites to the active center of pol II. In addition, YY1, TFIIB, and pol II must together be able to restrict initiation in only one direction from these adjacent start sites. If YY1-dependent transcription is not a simple "hit-and-run" process in which YY1 and TFIIB position pol II for a single round of transcription, additional requirements obtain. YY1 would have to remain in the vicinity of the promoter during separation of the template and nontemplate strands, and during formation of a DNA/RNA hybrid containing the template strand and the newly synthesized 5' end of the nascent RNA transcript. Once pol II begins to elongate and moves away from the start site, YY1 and the P5 Initiator element would then have to reestablish the YY1-DNA complex that supports specific, unidirectional initiation of transcription.

Structural Overview. The cocrystal structure of Δ YY1 complexed with a 20-bp oligonucleotide, corresponding to the AAV P5 initiator element, was obtained at 2.5-Å resolution (see Figs. 1–3 and Table 1). The Δ YY1-initiator element complex demonstrates many features of DNA recognition by zinc-finger proteins. All four fingers bind the major groove, using residues in similar positions within the finger to those responsible for stabilizing the Zif268-, GLI-, and Tramtrack-DNA complexes (16, 18, 21) (Fig. 1A). The duplex oligonucleotide most closely resembles B-form DNA, with conformational parameters typical of zinc finger-DNA cocrystal structures (22) (data not shown). Despite earlier electrophoretic findings (23), Δ YY1 does not bend DNA significantly, which

is also true of Zif268, GLI, and Tramtrack. This discrepancy is probably not due to YY1 truncation because the DNA-binding properties of Δ YY1 and YY1 are indistinguishable (see Fig. 4A). Our cocrystal structure also explains the DNA footprints of Δ YY1 and YY1, which are identical (A.U. and T.S., unpublished observations). Other transcription factor-DNA cocrystal structures have revealed straight DNA, where DNA bending was expected (e.g., Max; ref. 24).

Δ YY1-DNA Interactions and Initiator Element Recognition. Fig. 2B provides a schematic view of Δ YY1-DNA contacts. Three of Δ YY1's four zinc fingers make multiple contacts with major groove edges of bases (finger 1 makes a single base contact via Lys-315). The observed side chain-base contacts explain the YY1 binding consensus [5'-(C/g/a)(G/t)(C/t/a)CATN(T/a)(T/g/c)-3'] (26) and the functional consequences of changing individual bases within YY1 binding sites (27). The distribution of protein-DNA interactions between the template and nontemplate strands of the Initiator element changes abruptly at the two AAV P5 promoter transcription start sites specified by YY1 (Ade-10-Thy-31 and Thy-11-Ade-30, see Fig. 2B). These two positions represent the 5' most sites on the nontemplate strand that appear capable of tolerating DNA strand separation, without sacrificing stabilizing contacts between Δ YY1 and the DNA (Figs. 2 and 3). Upstream contacts are nearly evenly distributed between the two strands (template strand, 3/7 phosphoribose groups and 3/7 bases; nontemplate strand, 5/7 phosphoribose groups and 3/7 bases; see Fig. 2B). In contrast, between Ade-10 and Ade-17, there is a single direct protein-DNA interaction

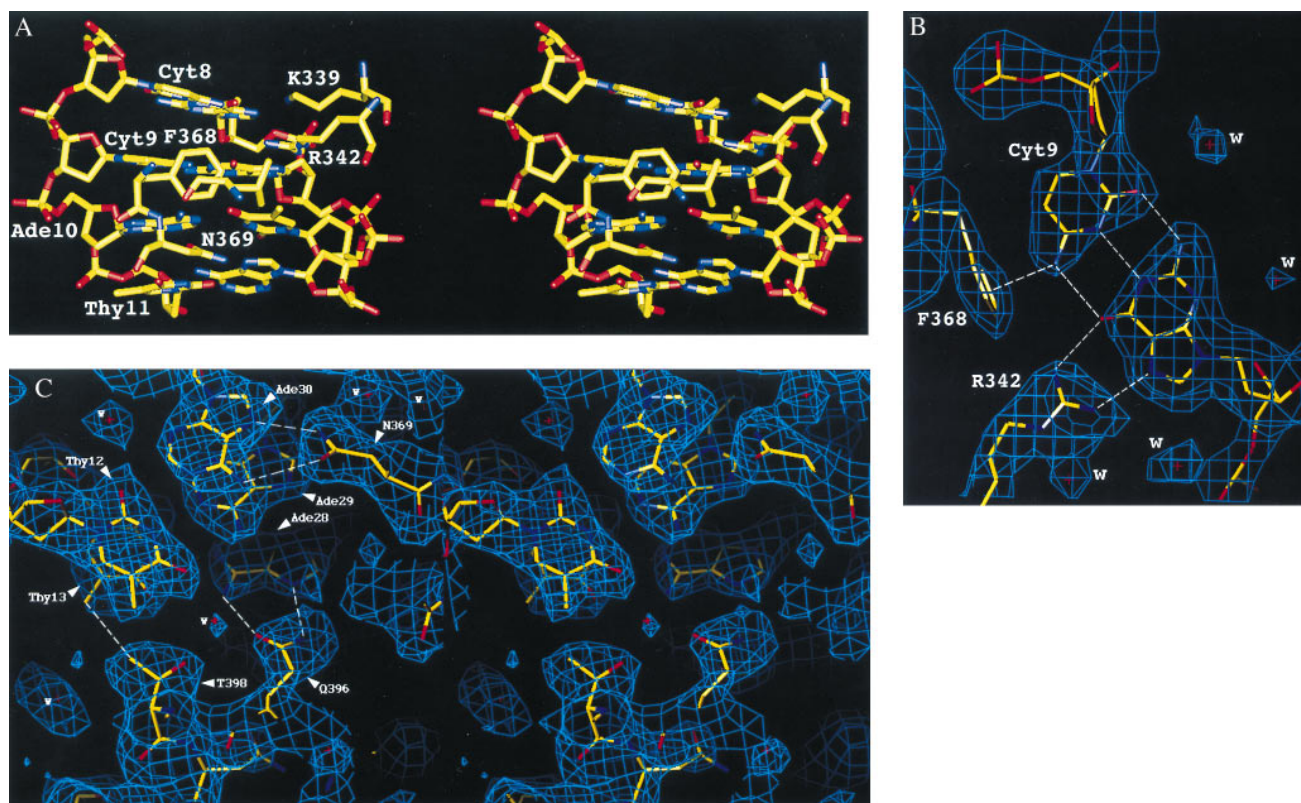


FIG. 3. DNA-protein contacts. (A) Stereoview of the Δ YY1-DNA complex in the vicinity of the transcription start sites, with the final atomic model drawn as a color-coded stick figure (carbon = yellow, oxygen = red, nitrogen = blue, and phosphorous = yellow). This view illustrates the transition from protein-DNA contacts distributed between both strands and segregation of contacts to the template strand from Ade-10-Thy-31 onwards. (B) Drawing of the final $(2|F_{\text{observed}}| - |F_{\text{calculated}}|)$ electron density map (contour level 1.0σ), showing the Cyt-9-Gua-32 bp interacting with Arg-342 and Phe-362. The NH \rightarrow Phe interaction between Cyt-9 and Phe-368 is also shown with a dashed line. N-H $\cdots\pi$ -electron cloud interactions have been characterized by Levitt and Perutz (28), and reviewed in ref. 29. (C) Stereoview of the final $(2|F_{\text{observed}}| - |F_{\text{calculated}}|)$ electron density map (contour level 1.0σ), showing the portion of Δ YY1-DNA complex downstream of the transcription start sites. Multiple side chain-base interactions occur between Δ YY1 and the template strand. Interactions with the nontemplate strand include a single contact between Thr-398 and Thy-13, and two putative water-mediated contacts involving a water molecule modeled in the weak electron density feature located between Thr-398 and Gln-396.

involving Thr-398 and Thy-13 of the nontemplate strand ($\text{CH}_3\text{-CH}_3 = 3.9 \text{ \AA}$). The remaining direct interactions between ΔYY1 and the DNA backbone and bases are restricted to the template strand (6/8 phosphoribose groups and 4/8 bases; see Fig. 2*B*).

ΔYY1 Binding to a DNA/RNA Hybrid. Zinc-finger proteins that interact with only one strand within the DNA duplex (e.g., Zif268; ref. 18) also recognize and bind tightly to DNA/RNA hybrids in which the DNA strand preserves the protein-DNA contacts (30). Between Ade-10 and Gua-20 on the nontemplate strand of the P5 initiator element, there are no ΔYY1 atoms within 4 \AA of C2' carbons that would create bad steric clashes with 2' hydroxyls of ribose groups. Thus, our cocrystal structure predicts that ΔYY1 will recognize a DNA/RNA hybrid in which the nontemplate strand downstream of the transcription start sites has been replaced by a complementary RNA segment. Fig. 4*B* illustrates the results of an electrophoretic mobility shift study of ΔYY1 interacting with various DNA/RNA hybrids. As expected, the protein *only* demonstrates wild-type binding affinity for the oligonucleotide in which the nontemplate strand downstream of the transcription start sites is RNA (Fig. 4*B*, compare lanes 10 and 14). The sole direct side chain-base contact observed between ΔYY1 and

the nontemplate strand of the P5 initiator element downstream of the start sites is lost in this DNA/RNA hybrid, because uridine replaces Thy-13 eliminating the C5 methyl group making van der Waals contact with Thr-398. The remaining DNA/RNA hybrid oligonucleotides display either no detectable binding [all ribose template strand (Fig. 4*B*, lane 4)] or significantly weakened binding [downstream ribose template strand (Fig. 4*B*, lane 6), upstream ribose template strand (Fig. 4*B*, lane 8), all ribose nontemplate strand (Fig. 4*B*, lane 12), upstream ribose nontemplate strand (Fig. 4*B*, lane 16)].

Model for YY1-Dependent Transcription Initiation. Fig. 2*D* contrasts the first steps of YY1-dependent versus TBP-dependent transcription initiation. TBP recognizes and deforms the TATA element, creating a tight, long-lived protein-nucleic acid complex (reviewed in ref. 31). In turn, TFIIB recognizes the preformed TBP-DNA complex and acts as a precise spacer or bridge between the TATA box and pol II, fixing the transcription start site (reviewed in ref. 13). The ΔYY1 -DNA complex does not demonstrate an analogous deformed DNA-protein assembly on which to recruit TFIIB and target pol II to the transcription start site. Since neither TFIIB nor pol II act via sequence-specific DNA binding, we must look within the ΔYY1 -DNA complex for an explanation of specific and unidirectional transcription initiation by the YY1-TFIIB-pol II complex.

The asymmetric distribution of protein-DNA interactions between the template and nontemplate strands of the Initiator element relative to the two transcription start sites suggests an elegant mechanistic model that explains specific, unidirectional YY1-dependent transcription initiation. Within the AAV P5 initiator the 5' end of the mRNA maps to the middle of the DNA recognized by YY1, a site which must be accessible to pol II during transcription initiation. In the absence of any knowledge of the structure of the YY1-initiator 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, our structure (Fig. 2*B*) shows that ΔYY1 interacts with both DNA strands upstream of the two base pairs where mRNA synthesis is initiated (Ade-10-Thy-31 and Thy-11-Ade-30), but interacts almost exclusively with the template strand downstream of these two base pairs. In fact, downstream of the two start sites the nontemplate strand is almost completely exposed to solvent (Fig. 2), 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 Ade-10 or Thy-11 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 Initiator element. Compelling, albeit indirect, support for this assertion comes from the results of studies of the AAV P5 promoter containing bubble mismatches (11). 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 bubble to facilitate DNA strand separation. The proposed model is further substantiated by the results of our gel mobility shift studies that demonstrate specific binding of ΔYY1 to a hybrid oligonucleotide in which the part of DNA that corresponds to the nascent transcript had been replaced with RNA. The complex between ΔYY1 and the DNA/RNA hybrid oligonucleotide represents an analog of the ternary DNA/RNA/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

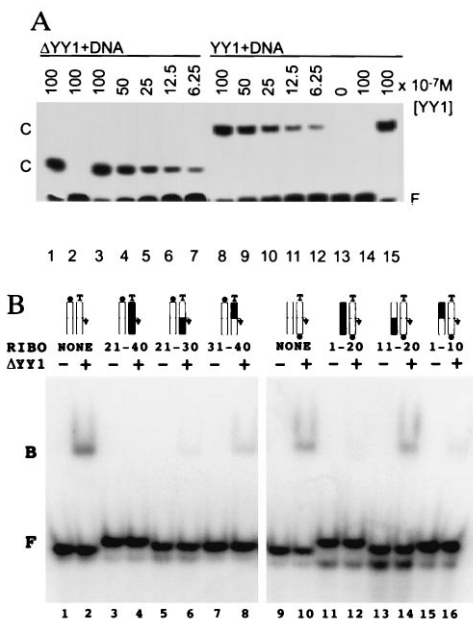


FIG. 4. Nucleic acid binding properties of YY1 and ΔYY1 . (*A*) YY1 and ΔYY1 have the same apparent affinity for the P5 initiator element, as judged by gel mobility shift assays. Reactions contained the same amount of radiolabeled crystallization DNA. The amount of YY1 and ΔYY1 added to each reaction is given above the lanes, and the positions of the protein-DNA complexes (C) are indicated. Lanes 2 and 14 include nonradioactive specific DNA competitor, and lanes 1 and 15 include nonradioactive nonspecific DNA competitor. (*B*) Binding of ΔYY1 to DNA/RNA hybrid analogs of the AAV P5 initiator element. The double-stranded synthetic oligonucleotides, radiolabeled at the 5' end of the template (lanes 1-8) or nontemplate (lanes 9-16) strand were incubated with (even lanes) or without (odd lanes) ΔYY1 and analyzed as in *A*. The ribonucleotide substituted regions and schematic representations of the double-stranded oligonucleotides are shown above the lanes (RNA is shown in black with the dominant start site denoted by an arrow, template strand is labeled "T" and the radioactive phosphate is represented by a black dot). The positions of the free probes (F) and the protein-nucleic acid complexes (B) are indicated. Note: Direct comparisons of the binding reactions containing probe labeled on the nontemplate strand (lanes 1-8) to binding reactions containing probe labeled on the template strand (lanes 9-16) are not possible because the probes had different specific radioactivity. Each binding reaction contains equivalent amounts of ΔYY1 (0.5 pmol) and radioactive probe (0.1 pmol).

been synthesized. During this process the nontemplate strand would presumably be sequestered by interactions with pol II.

One obvious question raised by this model concerns other zinc-finger proteins. Are they also likely to behave like YY1? Neither Zif268 nor GLI demonstrate the same abrupt transition in the middle of their binding sites (16, 18) (Fig. 1B) and neither protein has been shown to recognize pyrimidine-rich initiator elements or direct transcription initiation from any promoter. However, our model does predict that other zinc-finger proteins may support TBP-independent transcription initiation, provided that they recognize Initiator elements in a manner analogous to YY1.

CONCLUSION

Our study of the Δ YY1-initiator element complex demonstrates structural features that satisfy the central requirement for YY1-dependent transcription initiation. The Δ YY1-DNA complex leaves the nontemplate strand exposed to solvent downstream of the transcription start sites, explaining both start site selection and directionality of transcription. Moreover, YY1 can bind both to a transcription bubble and to a DNA/RNA hybrid. Together these results suggest that YY1 may indeed support multiple rounds of transcription initiation from the AAV P5 promoter, and work is underway to test this hypothesis. In addition, the role of YY1-dependent transcription initiation from nonviral gene promoters is being examined. The structure of the Δ YY1-DNA complex also provides the first three-dimensional view of a eukaryotic transcription factor recognizing an initiator element, which may be directly relevant to studies of the DNA-binding properties of TFIID (32) and TFIID (reviewed in ref. 4).

We thank S. Ramakoti for help with x-ray measurements; E. D. Halay for DNA synthesis; Drs. B. Chait and S. Cohen for mass spectrometry at the National Resource for Mass Spectrometric Analysis of Biological Macromolecules; Dr. E. Lolis for DNA/RNA hybrids; Drs. K. L. Clark, A. R. Ferré-D'Amaré, J. L. Kim, D. B. Nikolov, and P. Penev for discussions; A. Gazes for technical support; and Dr. M. Capel for help using the X12B Beamline. For useful suggestions, we are grateful to Drs. J. Kuriyan, N. Pavletich, and G. A. Petsko. H.B.H. is a Rockefeller University Graduate Fellow. This work was supported by a grant from the National Cancer Institute (T.S.) and by the Howard Hughes Medical Institute (S.K.B. and T.S.).

1. Bucher, P. & Trifonov, E. (1986) *Nucleic Acids Res.* **14**, 10009–10027.

2. Wobbe, C. & Struhl, K. (1990) *Mol. Cell. Biol.* **10**, 3859–3867.
3. Smale, S. T. & Baltimore, D. (1989) *Cell* **57**, 103–113.
4. Burley, S. K. & Roeder, R. G. (1996) *Annu. Rev. Biochem.* **65**, 769–799.
5. Roeder, R. G. (1991) *Trends Biochem. Sci.* **16**, 402–408.
6. Zawel, L., Kumar, K. & Reinberg, D. (1995) *Genes Dev.* **9**, 1479–1490.
7. Hahn, S. (1992) *Curr. Biol.* **2**, 152–154.
8. Seto, E., Shi, Y. & Shenk, T. (1991) *Nature (London)* **354**, 241–245.
9. Gill, G. (1994) *Curr. Biol.* **4**, 374–376.
10. Usheva, A. & Shenk, T. (1994) *Cell* **76**, 1115–1121.
11. Usheva, A. & Shenk, T. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13577–13582.
12. Parvin, J. & Sharp, P. (1993) *Cell* **73**, 533–540.
13. Nikolov, D. B., Chen, H., Halay, E., Usheva, A., Hisatake, K., Lee, D., Roeder, R. G. & Burley, S. K. (1995) *Nature (London)* **377**, 119–128.
14. Shi, Y., Seto, E., Chang, L.-S. & Shenk, T. (1991) *Cell* **67**, 377–388.
15. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
16. Pavletich, N. P. & Pabo, C. O. (1993) *Science* **261**, 1701–1707.
17. Terwilliger, T. C. & Eisenberg, D. (1983) *Acta Crystallogr. A* **39**, 813–817.
18. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
19. Brünger, A. T. (1992) *x-PLOR Version 3.1 Manual* (Yale Univ., New Haven, CT).
20. Laskowski, R. J., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–290.
21. Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366**, 483–487.
22. Nekludova, L. & Pabo, C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6948–6952.
23. Natesan, S. & Gilman, M. (1993) *Genes Dev.* **7**, 2497–2509.
24. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. (1993) *Nature (London)* **363**, 38–45.
25. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993) *Nature (London)* **365**, 520–527.
26. Hyde-DeRuyscher, R., Jennings, E. & Shenk, T. (1995) *Nucleic Acids Res.* **23**, 4457–4465.
27. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. & Smale, S. T. (1994) *Mol. Cell. Biol.* **14**, 116–127.
28. Levitt, M. & Perutz, M. (1988) *J. Mol. Biol.* **201**, 751–754.
29. Burley, S. K. & Petsko, G. A. (1988) *Adv. Protein Chem.* **39**, 125–192.
30. Shi, Y. & Berg, J. (1995) *Science* **268**, 282–284.
31. Burley, S. K. (1996) *Curr. Opin. Struct. Biol.* **6**, 69–75.
32. Roy, A. L., Malik, S., Meisterernst, M. & Roeder, R. G. (1993) *Nature (London)* **365**, 355–359.