Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 1119–1124, February 1996
Biochemistry

Protein–protein interactions in eukaryotic transcription initiation: Structure of the preinitiation complex

(TATA-element binding protein/RNA polymerase II/general transcription factors/alanine scanning)

HONG TANG*†, XIAOQING SUN†‡, DANNY REINBERG‡, AND RICHARD H. EBRIGHT*§

*Department of Chemistry and Waksman Institute, Rutgers University, New Brunswick, NJ 08855; and ‡Howard Hughes Medical Institute and Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, New Brunswick, NJ 08854

Communicated by Mark Ptashne, Harvard University, Cambridge, MA, November 6, 1995 (received for review September 20, 1995)

ABSTRACT **We have used alanine scanning to analyze protein–protein interactions by human TATA-element binding protein (TBP) within the transcription preinitiation complex. The results indicate that TBP interacts with RNA polymerase II and general transcription factors IIA, IIB, and IIF within the functional transcription preinitiation complex and define the determinants of TBP for each of these interactions. The results permit construction of a model for the structure of the preinitiation complex.**

Transcription initiation at eukaryotic protein-encoding genes is preceded by the assembly on promoter DNA of a preinitiation complex consisting of RNA polymerase II (Pol) and six general transcription factors (GTFs): i.e., IIA, IIB, IID, IIE, IIF, and IIH (1, 2). The human preinitiation complex contains at least 35 distinct polypeptide chains (at least 10 in Pol and at least 25 in the six GTFs) and has a molecular mass in excess of 2 MDa. Understanding human transcription initiation and transcription regulation will require elucidation of the arrangement of these numerous polypeptide chains relative to promoter DNA and relative to each other.

The first step in assembly of the preinitiation complex is binding of IID to a specific DNA sequence—the TATA element—located upstream of the transcription start (1, 2). IID is a multisubunit factor, consisting of the polypeptide chain responsible for recognition of the TATA element, termed TATA-element binding protein (TBP), and at least eight additional polypeptide chains, termed TBP-associated factors (TAFs) (1–4). TBP, by itself, is able to bind to promoter DNA in a sequence-specific fashion and to nucleate subsequent stepwise association of IIA, IIB, IIF-Pol, IIE, and IIH, yielding a preinitiation complex fully functional in basal transcription initiation (albeit not functional in activator-dependent transcription initiation.) TBP is thought to contain determinants for protein–protein interactions with Pol, GTFs, TAFs, and a number of transcription activators and repressors.

Within the past 2 years, the crystallographic structure of the yeast TBP–DNA complex has been determined to 2.5 Å resolution, and the crystallographic structure of the *Arabidopsis thaliana* TBP–DNA complex has been determined to 1.9 Å resolution (5, 6). The conserved DNA-binding, GTF-binding, TAF-binding core domain of TBP consists of two 80-amino acid imperfect direct repeats. Each repeat consists of two α-helices and five β-strands in the order β_1–α_1–β_2–β_3–β_4–β_5–α_2. The TBP–DNA complex is approximately twofold symmetric, with the first repeat responsible for recognition of the 3′ half of the TATA element and with the second repeat responsible for recognition of the 5′ half of the TATA element. TBP–DNA interactions are mediated through contacts between β_1, β_2, β_3, β_4, and β_5 of the first and second repeats and the DNA minor groove. TBP sharply bends DNA in the TBP–DNA complex.

The availability of the crystallographic structure of the TBP–DNA complex makes possible a systematic structure–function analysis of TBP–Pol, TBP–GTF, TBP–TAF, TBP–activator, and TBP–repressor interactions.

MATERIALS AND METHODS

Plasmids Encoding TBP Derivatives. Plasmid pHTT7f1-NH-TBP encodes 7 nonnative amino acids (MHHHHHH) followed by amino acids 2–339 of human TBP under control of the bacteriophage T7 gene 10 promoter. Plasmid pHTT7f1-NH-TBP was constructed by replacement of the *Eco*RI/*Hin*dIII segment of plasmid pET21(+) (Novagen) by the *Eco*RI/*Hin*dIII segment of plasmid pTK590 (obtained from T. Kerppola and T. Curran). Site-directed mutagenesis (7) was used to construct plasmids encoding alanine-substituted TBP derivatives.

Preparation of TBP Derivatives. TBP derivatives were overproduced in transformants of *Escherichia coli* strain BL21(DE3) (Novagen), purified under denaturing conditions by metal ion affinity chromatography on Ni^{2+}-NTA-agarose (Qiagen, Chatsworth, CA; procedure of the manufacturer, but with pH 5.75 in the third wash buffer), dialyzed 4 h at 4°C against two changes of 2000 vol of 8 M urea/100 mM sodium phosphate, pH 7.9/10 mM Tris·HCl/10 mM $MgCl_2$/2 mM dithiothreitol/1 mM EDTA/0.2 mM EGTA/0.2 mM phenylmethylsulfonyl fluoride/0.1% Nonidet P-40 (NP-40), dialyzed 15 h at 4°C against two changes of 4000 vol of 20 mM Tris·HCl, pH 7.9/100 mM KCl/10 mM $MgCl_2$/2 mM dithiothreitol/1 mM EDTA/0.2 mM EGTA/0.2 mM phenylmethylsulfonyl fluoride/0.1% NP-40/10% glycerol, and desalted into 20 mM Hepes·NaOH, pH 7.9/100 mM KCl/1 mM dithiothreitol/0.1 mM EDTA/0.2 mM phenylmethylsulfonyl fluoride/10% glycerol on Sephadex G-25 DNA grade (Pharmacia). Typically, the yield was 0.5 mg of TBP derivative per liter of bacterial culture. Protein preparations were stored in aliquots at −80°C.

Specific activities of protein preparations were normalized by analysis of TBP–DNA interaction. Reaction mixtures contained (in 20 μl): 0.3 nM ^{32}P-labeled DNA fragment containing positions −40 to +15 of the adenovirus major late promoter (25 Bq/fmol) and 0–200 nM TBP derivative in buffer A [20 mM Tris·HCl, pH 7.9/20 mM Hepes·NaOH/60 mM KCl/10 mM $MgCl_2$/8 mM $(NH_4)_2SO_4$/0.5 mM dithiothreitol/0.05 mM EDTA/0.1 mM phenylmethylsulfonyl fluoride/25 μg of poly(dG-dC) per ml (average size, 700 kDa)/25 mg of polyethylene glycol per ml (average size, 8 kDa)/5% glycerol]. Reaction mixtures were incubated 30 min at 30°C. Reaction

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.

Abbreviations: Pol, RNA polymerase II; GTF, general transcription factor; TBP, TATA-element binding protein; TAF, TBP-associated factor.

†H.T. and X.S. contributed equally to this work.

§To whom reprint requests should be addressed at: Waksman Institute, Rutgers University, New Brunswick, NJ 08855.

products were analyzed by electrophoresis in 45 mM Tris borate, pH 8.0/10 mM $MgCl_2$/0.1 mM EDTA through 5% polyacrylamide (37.5:1 acrylamide/bisacrylamide)/2.7% glycerol slab gels (9 × 7.5 × 0.15 cm; 20 V/cm; 30 min at 23°C), followed by PhosphorImaging. Apparent binding constants were calculated by nonlinear regression (equation in ref. 8). TBP derivatives having alanine substitutions at amino acids 283 and 294 were severely defective in TBP–DNA complex formation. In subsequent experiments, these TBP derivatives were omitted. In subsequent experiments, concentrations of the remaining TBP derivatives were adjusted to levels that yield 25% saturation of TBP–DNA complex formation.

Analysis of TBP Derivatives: Complex Formation. Reaction mixtures contained (in 20 μl): 0.3 nM ^{33}P-labeled DNA fragment containing positions −40 to +15 of the adenovirus major late promoter (10 Bq/fmol), TBP derivative, and, where indicated, human IIA (S200 fraction; ref. 9), recombinant human IIB (10), recombinant human IIF (11), human Pol (DEAE-5PW fraction; ref. 12), recombinant human IIE (13), and human IIH (0.5 M phenyl-Superose fraction; ref. 14) in buffer A. Where indicated as fixed, concentrations of IIA, IIB, IIF, Pol, and IIE were 10, 8, 30, 50, and 60 nM, respectively. Reaction mixtures were prepared in stepwise fashion, with incubation 30 min at 30°C after addition of each successive factor. Reaction products were analyzed by electrophoresis as described above (Figs. 1*A* and 2*A*) or by electrophoresis in 45 mM Tris borate, pH 8.0/0.1 mM EDTA through 5% polyacrylamide (37.5:1, acrylamide/bisacrylamide)/2.7% glycerol slab gels [27 × 15 × 0.15 cm; 7 V/cm; 2 h at 23°C (Figs. 1*B* and 2*B*); 3 h at 23°C (Figs. 1 *C* and *D* and 2 *C* and *D*), 4 h at 23°C (Figs. 1 *E* and 2*E*), or 8 h at 23°C (Figs. 1*F* and 2*F*)], followed by PhosphorImaging. Binding constants were calculated by nonlinear regression (equation in ref. 8).

TBP derivatives having alanine substitutions at amino acids 284, 286, and 287 did not support saturation of TBP–DNA–IIA–IIB complex formation and therefore were not assayed for interactions with IIF, Pol, IIE, and IIH.

Analysis of TBP Derivatives: Transcription. Reaction mixtures contained (in 40 μl): 3 nM plasmid pML(C_2AT)Δ-50 DNA (which carries positions −50 to +10 of the adenovirus major late promoter followed by a 382 nt of G:C-less cassette; ref. 23), TBP derivative, human IIA (S200 fraction; ref. 9), recombinant human IIB (10), recombinant human IIF (11), human Pol (DEAE-5PW fraction; ref. 12), recombinant human IIE (13), human IIH (0.5 M phenyl-Superose fraction; ref. 14), 15 μM [α-^{32}P]UTP (1.5 Bq/fmol), 0.6 mM ATP, 0.6 mM CTP, and 20 units of RNase T1 in 20 mM Hepes·NaOH, pH 7.9/50 mM KCl/4 mM $MgCl_2$/10 mM $(NH_4)_2SO_4$/10 mM 2-mercaptoethanol/20 mg of polyethylene glycol per ml (average size, 8 kDa)/12% glycerol. Where fixed, concentrations of IIA, IIB, IIF, Pol, IIE, and IIH were 25, 20, 20, 30, 20, and 50 nM, respectively. Where variable, concentrations of IIB were 0.4, 2, 10, and 50 nM; concentrations of IIF were 1.6, 8, and 40 nM; and concentrations of Pol were 0.32, 1.6, 8, and 40 nM. Reaction components except NTPs were incubated 20 min at 30°C. Reactions were initiated by addition of NTPs and were terminated after 1 h at 30°C by addition of 80 μl 100 mM sodium acetate, pH 5.5/10 mM EDTA/0.2% SDS/1 mg of yeast tRNA per ml. Products were phenol/chloroform extracted, ethanol precipitated, and analyzed by urea/PAGE, followed by autoradiography and PhosphorImaging.

Molecular Modeling. The structure of the human TBP–DNA complex was homology modeled based on the crystallographic structure of the yeast TBP–DNA complex (ref. 5; atomic coordinates obtained from P. Sigler). Amino acid sequences of human and yeast TBP core domains are 81% identical with no gaps or insertions (1–3). Amino acids of yeast TBP core domain not identical in human TBP core domain were replaced, retaining yeast χ_1 side-chain torsion angles and using most favored χ_n side-chain torsion angles. In a small

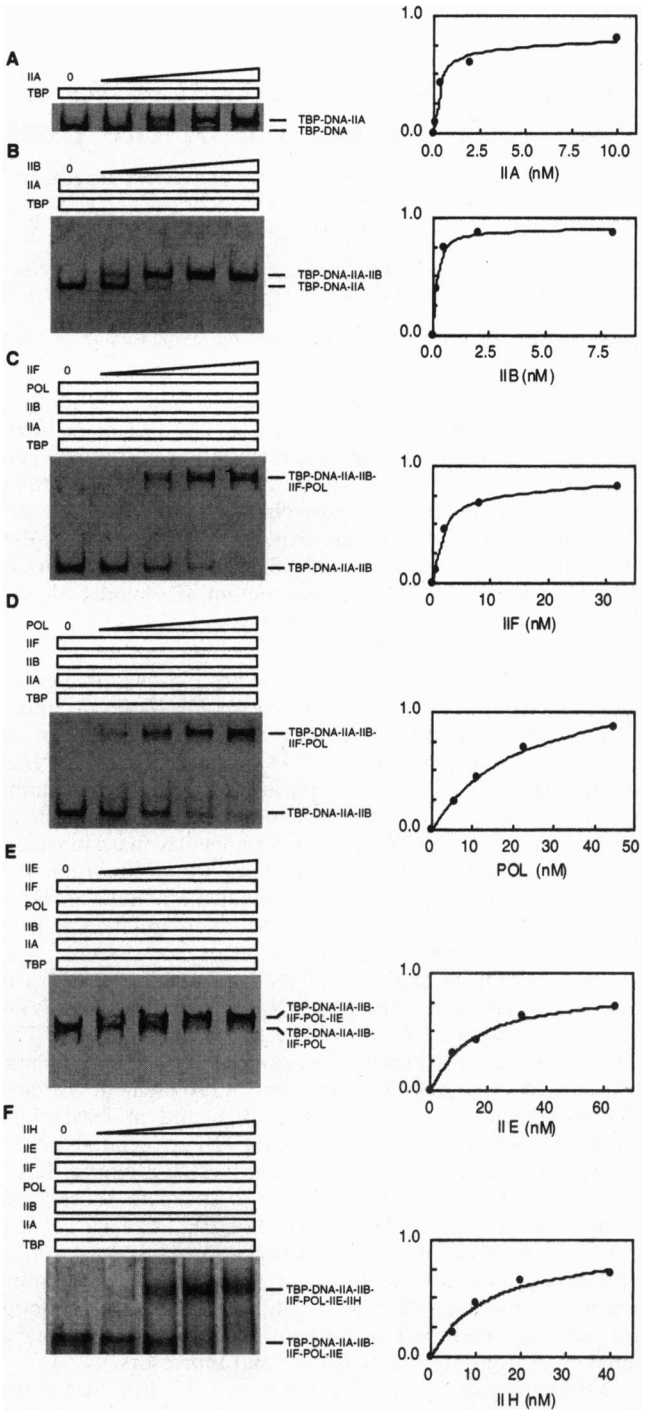

FIG. 1. Interactions of wild-type TBP with IIA (*A*), IIB (*B*), IIF (*C*), Pol (*D*), IIE (*E*), and IIH (*F*). (*Left*) Autoradiograms. (*Right*) Calculated binding curves.

number of cases, it was necessary to change χ_1 of the replaced amino acid to avoid steric clash. Human TBP core domain contains an additional C-terminal amino acid not present in yeast TBP core domain (amino acid 339); this amino acid was modeled using an extended backbone conformation and most favored side-chain torsion angles.

The accessible surface of TBP in the homology-modeled human TBP–DNA complex was calculated using the algorithm of ref. 24. Calculations were performed using a probe of radius 4 Å—a probe comparable in size to an amino acid side chain and, thus, comparable in size to a potential point of contact with Pol or a GTF.

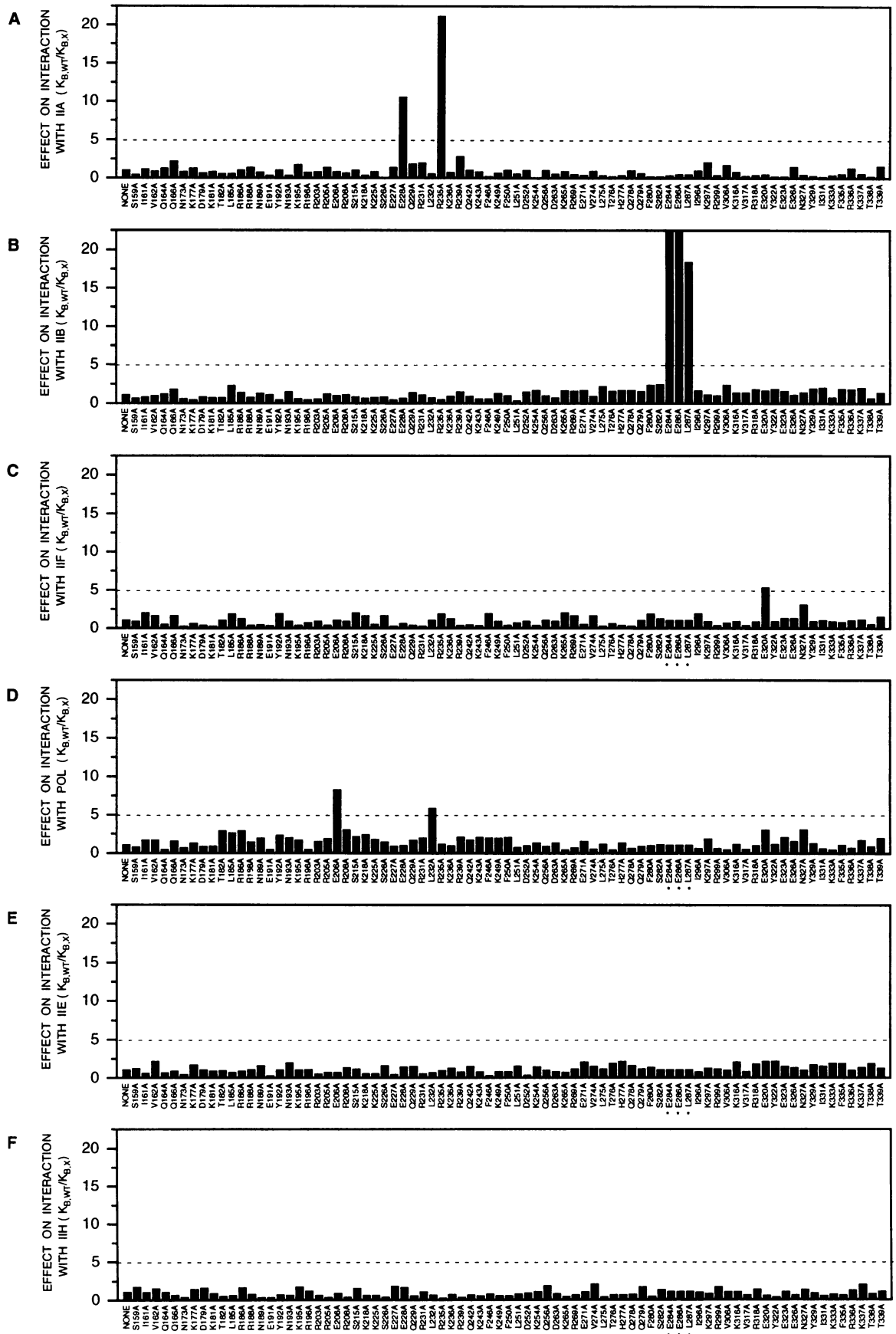

FIG. 2. Interactions of wild-type TBP and alanine-substituted TBP derivatives with IIA (*A*), IIB (*B*), IIF (*C*), Pol (*D*), IIE (*E*), and IIH (*F*). In each panel, a dashed line is drawn at the level of a 5-fold defect in equilibrium binding constant. A 5-fold defect in equilibrium binding constant corresponds to a difference in binding free energy of ≈1 kcal/mol—a value comparable to the free energy contribution of a single side-chain interaction (15, 16). Alanine substitutions that result in ≥5-fold defects in equilibrium binding constant define amino acids that are candidates to participate in direct, energetically favorable, side-chain interactions (see refs. 17–22).

Modeled B-DNA coordinates for the DNA segments upstream and downstream of the TATA element were generated using INSIGHT II (Biosym Technologies, San Diego). Radii of spheres corresponding to polypeptide chains were calculated using the equation $r_{Å} = (0.29M)^{1/3}$, where M is the molecular mass (see ref. 25).

RESULTS AND DISCUSSION

Alanine scanning—systematic construction of single-alanine substitutions and determination of effects on function—permits identification of individual amino acid side-chain determinants for protein–protein interaction (17–22). Alanine scanning has two key advantages. First, alanine scanning yields a comprehensive set of substitutions, including both phenotypically positive and phenotypically negative substitutions. Second, and more important, alanine scanning yields side-chain truncation substitutions; alanine substitution eliminates all side-chain atoms beyond C_β.

In this work, we have used alanine scanning to analyze TBP–Pol and TBP–GTF interactions. Molecular modeling indicates that 81 nonproline amino acids of human TBP core domain have side-chain atoms beyond C_β accessible on the surface of the human TBP–DNA complex. We reason that these 81 amino acids constitute the entire set—or nearly the entire set—of amino acids of TBP that are candidates to make side-chain interactions with Pol and GTFs. For each of these 81 amino acids, we have constructed a single alanine substitution. We then have quantified the abilities of wild-type TBP and alanine-substituted TBP derivatives to interact successively with IIA, IIB, IIF, Pol, IIE, and IIH (Figs. 1 and 2).

Interaction with IIA. Alanine substitution of two—and only two—amino acids within human TBP core domain resulted in a ≥5-fold reduction in equilibrium binding constant for interaction with IIA; i.e., Glu-228 and Arg-235 (Fig. 2*A*). We conclude that for these two amino acids—and for no other amino acids of human TBP core domain—side-chain atoms beyond C_β are critical for interaction with IIA. In the structure of the TBP–DNA complex, amino acids 228 and 235 are located adjacent to each other on the same face of α2 of the first repeat of TBP and form a surface with dimensions of ≈6 Å × ≈18 Å (Fig. 3*A*). We propose that amino acids 228 and 235 make direct contact with IIA in the TBP–DNA–IIA complex.

Our results confirm and extend reports that substitutions in and following α2 of the first repeat of yeast TBP affect interaction with IIA (26, 27).

Interaction with IIB. Alanine substitution of three—and only three—amino acids within human TBP core domain resulted in a ≥5-fold reduction in equilibrium binding constant for interaction with IIB; i.e., Glu-284, Glu-286, and Leu-287 (Fig. 2*B*). *In vitro* transcription experiments confirm that TBP derivatives having alanine substitutions at amino acid 284, amino acid 286, or amino acid 287 are defective in transcription initiation and that the defects can be overcome by excess IIB (Fig. 4*A*). We conclude that for these three amino acids—and for no other amino acids of human TBP core domain—side-chain atoms beyond C_β are critical for interaction with IIB. In the structure of the TBP–DNA complex, amino acids 284, 286, and 287 are located immediately adjacent to each other in the β2–β3 loop (the "stirrup") of the second repeat of TBP and form a surface with dimensions of ≈11 Å × ≈16 Å (Fig. 3*B*). We propose that amino acids 284, 286, and 287 make direct contact with IIB in the TBP–DNA–IIA–IIB complex.

Our results confirm and extend reports that a substitution in the β2–β3 loop of the second repeat of yeast TBP (at the position equivalent to amino acid 287 of human TBP) affects interaction with IIB (28, 29).

Interaction with IIF. Alanine substitution of one—and only one—amino acid within human TBP core domain resulted in a ≥5-fold reduction in equilibrium binding constant for interaction with IIF; i.e., Glu-320 (Fig. 2*C*). The magnitude of the defect is modest, being only slightly greater than 5-fold. Nevertheless, the effect is real and reproducible. Furthermore, *in vitro* transcription experiments confirm that the TBP derivative having an alanine substitution at amino acid 320 is defective in transcription initiation (Fig. 4*B*). We conclude that for amino acid 320—and for no other amino acid of human TBP core domain—side-chain atoms beyond C_β are critical for interaction with IIF.

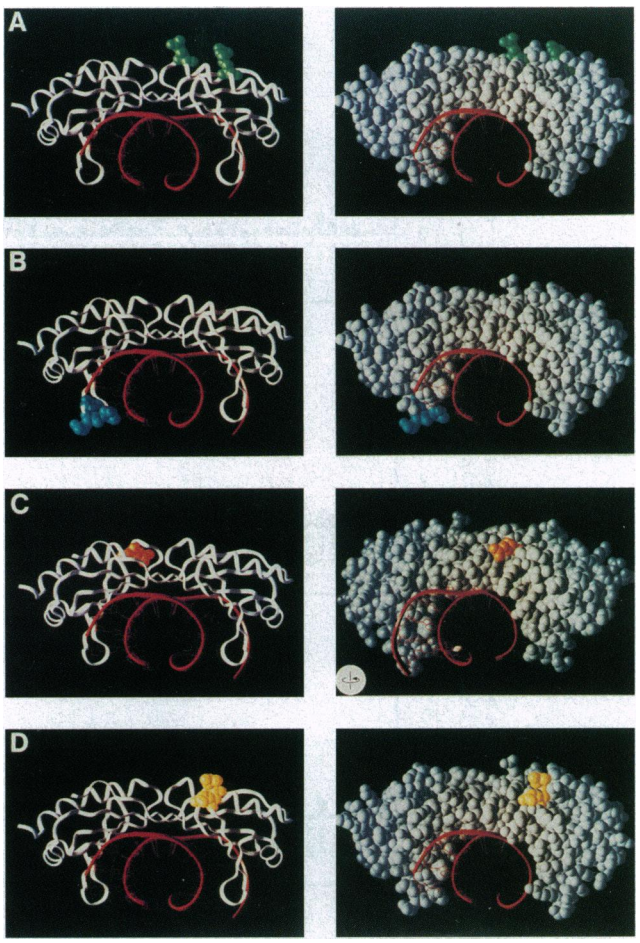

FIG. 3. Homology-modeled structure of the human TBP–DNA complex showing determinants for interaction with IIA (*A*), IIB (*B*), IIF (*C*), and Pol (*D*). (*Left*) Ribbon representations. (*Right*) van der Waals representations. The van der Waals representation in *C* is rotated 180° relative to the ribbon representations and other van der Waals representations.

In the structure of the TBP–DNA complex, amino acid 320 is located in α2 of the second repeat of TBP and forms a surface of ≈5 Å × ≈7 Å (Fig. 3*C*). We propose that amino acid 320 makes direct contact with IIF in the TBP–DNA–IIA–IIB–IIF–Pol complex.

Alanine substitution of amino acid 327 resulted in a modest, but reproducible, reduction in interaction with IIF (≈3-fold; Fig. 2*C*). Amino acid 327 is located adjacent to amino acid 320 on the same face of α2 of the second repeat and may constitute an additional, energetically less significant, point of contact with IIF.

Interaction with Pol. Alanine substitution of two—and only two—amino acids within human TBP core domain resulted in a ≥5-fold reduction in equilibrium binding constant for interaction with Pol; i.e., Glu-206 and Leu-232 (Fig. 2*D*). *In vitro* transcription experiments confirm that TBP derivatives having alanine substitutions at amino acid 206 or amino acid 232 are defective in transcription initiation (Fig. 4*C*). We conclude that

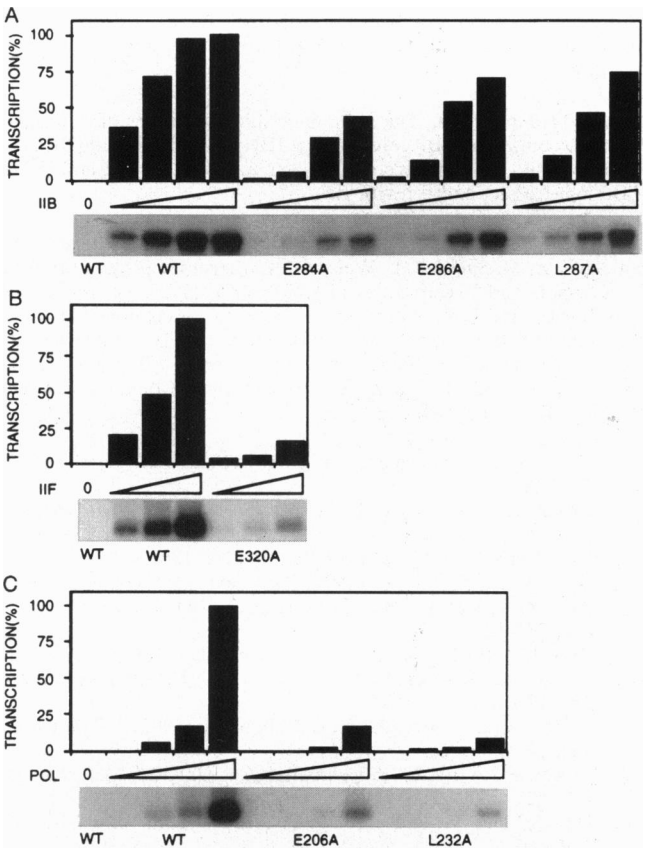

FIG. 4. Effects on transcription initiation of alanine substitutions within the determinants for interaction with IIB (*A*), IIF (*B*), and Pol (*C*). Autoradiograms show 392-nt transcripts initiated at the adenovirus major late promoter.

for these two amino acids—and for no other amino acids of human TBP core domain—side-chain atoms beyond C_β are critical for interaction with Pol.

In the structure of the TBP–DNA complex, amino acid 206 is located in the $\beta3$–$\beta4$ loop of the first repeat of TBP, and amino acid 232 is located in $\alpha2$ of the first repeat of TBP (Fig. 3*D*). Although these amino acids are distant in the primary structure, they are immediately adjacent in the three-dimensional structure and form a continuous surface with dimensions of ≈7 Å × ≈15 Å. We propose that amino acids 206 and 232 make direct contact with Pol in the TBP–DNA–IIA–IIB–IIF–Pol complex.

Alanine substitution of amino acids 182, 185, 186, and 208 resulted in modest, but reproducible, reductions in interaction with Pol (≈3-fold; Fig. 2*D*). Amino acids 182, 185, and 186 are located in $\alpha1$ of the first repeat, and amino acid 208 is located in the $\beta3$–$\beta4$ loop of the first repeat. These amino acids are near amino acids 206 and 232 and may constitute additional, energetically less significant, points of contact with Pol.

The determinant for interaction with Pol is adjacent to the determinant for interaction with IIA (Figs. 3 *A* and *D*). This raises the possibility that the determinant might reflect a Pol-dependent extension of the contact by IIA rather than a direct contact by Pol. However, this possibility is excluded by control experiments indicating that alanine substitution of amino acid 206 or 232 has similar effects on interaction with Pol in the presence of IIA (TBP–DNA–IIA–IIB–IIF–Pol complex formation) and in the absence of IIA (TBP–DNA–IIB–IIF–Pol complex formation).

Interaction with IIE and IIH. Alanine substitution of no amino acid of human TBP core domain resulted in a ≥5-fold reduction in interaction with IIE or IIH (Figs. 2 *E* and *F*). We conclude that for no amino acid of human TBP core domain are side-chain atoms beyond C_β critical for interaction with IIE or IIH.

We propose that human TBP core domain makes no direct interactions with IIE and IIH within the preinitiation complex. However, we emphasize that this proposal must be considered tentative, since our analysis excludes proline atoms, side-chain C_β atoms, and backbone atoms, and since our threshold for significance of effects excludes atoms involved in weak or neutral interactions [interactions contributing less than ≈1 kcal/mol binding free energy (see legend to Fig. 2)].

Structure of Preinitiation Complex. Our results permit construction of models of the successive higher-order complexes containing IIA, IIB, IIF, and Pol (Fig. 5).

IIA consists of three distinct polypeptide chains with molecular masses of 14, 19, and 34 kDa (1, 2). Our results indicate that IIA interacts with the upstream face of TBP (Fig. 5*A*). Published results indicate that IIA interacts with the DNA segment upstream of the TATA element (30). Therefore, we model IIA interacting simultaneously with the upstream face of TBP and the upstream DNA segment (Fig. 5*A*).

IIB consists of a single polypeptide chain with a molecular mass of 35 kDa (1, 2). The IIB core domain, which is sufficient for interaction with the TBP–promoter complex, consists of two 75-amino acid imperfect direct repeats (1, 2). Our results indicate that IIB interacts with the underside of TBP (Fig. 5*A*). Therefore, we model IIB core domain interacting with the underside of TBP (Fig. 5*B*). The determinant of TBP for interaction with IIB core domain is located within the minor groove of the DNA segment immediately upstream of the TATA element. We infer that, in the preinitiation complex,

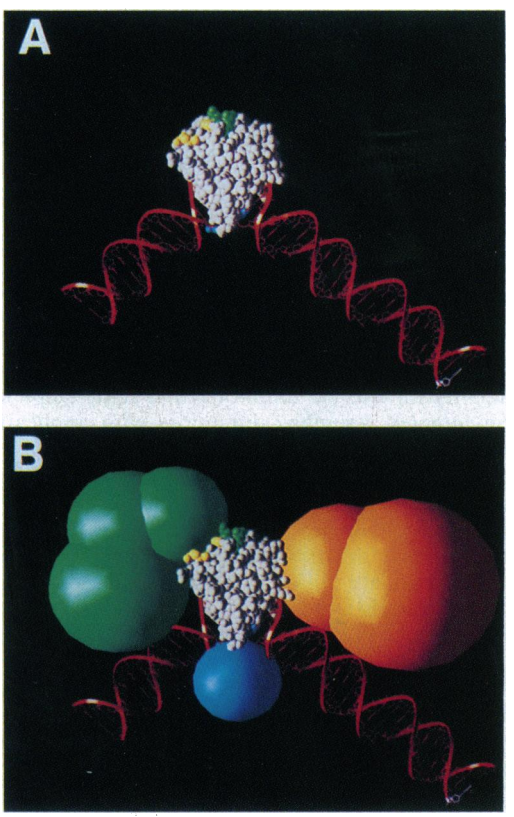

FIG. 5. (*A*) Model for structure of the complex of human TBP with the adenovirus major late promoter (see *Materials and Methods*; see also refs. 5 and 6). The transcription start site is white. The determinants for interaction with IIA, IIB, and Pol are green, blue, and yellow, respectively. The determinant for interaction with IIF is not visible in this orientation. (*B*) Model for structure of the higher-order complex containing IIA (green spheres), IIB (blue spheres), IIF (orange spheres), and Pol. Pol is envisioned as extending from TBP to the transcription start site.

residues of IIB core domain are located within the minor groove immediately upstream of the TATA element and, therefore, that the IIB core domain is likely to be a DNA binding protein. Given the proximity—due to the TBP-induced DNA bend—of the DNA segments upstream and downstream of the TATA element, and given the size of the IIB core domain, we infer further that IIB core domain is likely to interact simultaneously with the DNA segments upstream and downstream of the TATA element (Fig. 5*B*). The proposal that IIB core domain interacts with DNA segments upstream and downstream of the TATA element receives strong support from DNase I footprinting (31), hydroxyl-radical footprinting (32), and site-specific protein–DNA photocrosslinking (T. Lagrange, G. Orphanides, Y. Ebright, D.R., and R.H.E., unpublished data).

IIF consists of two distinct polypeptide chains with molecular masses of 30 and 74 kDa (1, 2). Our results indicate that IIF interacts with the downstream face of TBP (Fig. 5*A*). Published results indicate that IIF contains a DNA binding domain (33) and interacts with the DNA segment downstream of the TATA element (30). Therefore, we model IIF interacting simultaneously with the downstream face of TBP and the downstream DNA segment (Fig. 5*B*).

Pol consists of at least 10 distinct polypeptide chains and has a molecular mass in excess of 500 kDa (1, 2). Our results indicate that Pol interacts with the right flank of TBP (Fig. 5). Pol also must interact with the transcription start site. Therefore, we envision Pol as interacting simultaneously with the right flank of TBP and the transcription start site. According to this model, Pol would interact with a surface of TBP immediately adjacent to the determinant for interaction with IIA and therefore almost certainly would interact with IIA. In addition, Pol would extend past IIB and IIF and therefore potentially would interact with these factors (see refs. 1 and 2).

Protein–Protein Interactions Within the Preinitiation Complex. The determinants for protein–protein interaction defined in this work are small (40–200 Å^2) and contain small numbers of critical amino acids (1–3 critical amino acids) (Fig. 3). Both the small size of the determinants and the presence of only a small number of critical amino acid side chains in the determinants are reminiscent of patterns observed with functional epitopes for protein–antibody interaction, protein–receptor interaction, and prokaryotic transcription factor interactions (17–22). The small size of the determinants suggests that it may be possible to design low molecular weight mimics of the determinants for use as inhibitors of specific steps in transcription-complex assembly—inhibitors with applications as tools for molecular biology research and, potentially, with applications as therapeutic agents (see refs. 19 and 20).

Although the determinants for protein–protein interaction defined in this work are small, there are four interacting factors, and the interacting factors are large. Therefore, the fraction of the surface of the TBP core domain screened by the interacting factors is large. This is especially true with respect to the first repeat of TBP core domain. According to the model in Fig. 5*B*, virtually the entire upstream face of the first repeat is screened by IIA, virtually the entire downstream face of the first repeat is screened by IIF, and virtually the entire right flank of the first repeat is screened by Pol. We infer that other factors able to interact with the TBP core domain in the context of the preinitiation complex—such as TAFs—are unlikely to interact with the first repeat and are likely instead to interact with the second repeat (for which, according to the model, >60% of the surface remains available).

Prospect. Our results have permitted construction of a model for the structure of the preinitiation complex. This model provides a framework for incorporation of high-resolution structures of individual polypeptide chains as they become available from crystallographic and NMR spectroscopic studies, for understanding transcription initiation, and for designing low molecular weight inhibitors of transcription initiation.

Note Added in Proof. The crystallographic structure of a ternary complex containing TBP core domain, IIB core domain, and DNA has been reported (34). In this structure, amino acids 284, 286, and 287 of TBP make direct contact with IIB.

We thank R. Drapkin, P. Kumar, E. Maldonado, F. Mermelstein, R. Shiekhattar, M. Sheldon, L. Weis, and L. Zawel for protein samples; T. Kerppola and T. Curran for plasmid pTK590; P. Sigler for atomic coordinates; and L. Guarente and S. Hahn for discussion. This work was supported by National Institutes of Health (NIH) Grant GM37120 and a Howard Hughes Medical Institute Investigatorship to D.R. and by NIH Grant GM51527 and New Jersey Cancer Commission Grant 94-45-CCR-00 to R.H.E.

1. Conaway, R. & Conaway, J. (1993) *Annu. Rev. Biochem.* **62,** 161–190.
2. Zawel, L. & Reinberg, D. (1993) *Prog. Nucleic Acids Res.* **44,** 67–108.
3. Hernandez, N. (1993) *Genes Dev.* **7,** 1291–1308.
4. Goodrich, J. & Tjian, R. (1994) *Curr. Opin. Cell Biol.* **6,** 403–409.
5. Kim, Y., Geiger, J., Hahn, S. & Sigler, P. (1993) *Nature (London)* **365,** 512–520.
6. Kim, J. & Burley, S. (1994) *Nat. Struct. Biol.* **1,** 638–652.
7. Kunkel, T., Bebenek, K. & McClary, J. (1991) *Methods Enzymol.* **204,** 125–138.
8. Gunasekera, A., Ebright, Y. & Ebright, R. (1992) *J. Biol. Chem.* **267,** 14713–14720.
9. Cortes, P., Flores, O. & Reinberg, D. (1992) *Mol. Cell. Biol.* **12,** 413–421.
10. Ha, I., Lane, W. & Reinberg, D. (1991) *Nature (London)* **352,** 689–695.
11. Wang, B. Q., Kostrun, C., Finkelstein, A. & Burton, Z. (1993) *Protein Express. Purif.* **4,** 207–214.
12. Lu, H., Flores, O., Weinmann, R. & Reinberg, D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10004–10008.
13. Peterson, M. G., Inostroza, J., Maxon, M., Flores, O., Admon, A., Reinberg, D. & Tjian, R. (1991) *Nature (London)* **354,** 369–373.
14. Flores, O., Lu, H. & Reinberg, D. (1992) *J. Biol. Chem.* **267,** 2786–2793.
15. Nemethy, G. & Scheraga, H. (1962) *J. Phys. Chem.* **66,** 1773–1789.
16. Fersht, A., Shi, J.-P., Knill-Jones, J., Lowe, D., Wilkinson, A., Blow, D., Brick, P., Carter, P., Waye, M. & Winter, G. (1985) *Nature (London)* **314,** 235–238.
17. Cunningham, B. & Wells, J. (1989) *Science* **244,** 1081–1085.
18. Jin, L., Fendly, B. & Wells, J. (1992) *J. Mol. Biol.* **226,** 851–865.
19. Cunningham, B. & Wells, J. (1993) *J. Mol. Biol.* **234,** 554–563.
20. Clackson, T. & Wells, J. (1995) *Science* **267,** 383–386.
21. Kelley, R. & O'Connell, M. (1993) *Biochemistry* **32,** 6828–6835.
22. Niu, W., Zhou, Y., Dong, Q., Ebright, Y. & Ebright, R. (1994) *J. Mol. Biol.* **243,** 595–602.
23. Sawadogo, M. & Roeder, R. (1985) *Cell* **43,** 165–175.
24. Connolly, M. (1983) *Science* **221,** 709–713.
25. Cantor, C. & Schimmel, P. (1980) *Biophysical Chemistry, Part II* (Freeman, San Francisco), pp. 550–555.
26. Lee, D. K., Dejong, J., Hashimoto, S., Horikoshi, M. & Roeder, R. (1992) *Mol. Cell. Biol.* **12,** 5189–5196.
27. Buratowski, S. & Zhou, H. (1992) *Science* **255,** 1130–1132.
28. Kim, T. K., Hashimoto, S., Kelleher, R., III, Flanagan, P., Kornberg, R., Horikoshi, M. & Roeder, R. (1994) *Nature (London)* **369,** 252–255.
29. Kim, T. K., Zhao, Y., Ge, H., Bernstein, R. & Roeder, R. (1995) *J. Biol. Chem.* **270,** 10976–10981.
30. Coulombe, B., Li, J. & Greenblatt, J. (1994) *J. Biol. Chem.* **269,** 19962–19967.
31. Malik, S., Lee, D. K. & Roeder, R. (1993) *Mol. Cell. Biol.* **13,** 6253–6259.
32. Lee, S. & Hahn, S. (1995) *Nature (London)* **376,** 609–612.
33. Tan, S., Garrett, K. P., Conaway, R. & Conaway, J. W. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9808–9812.
34. Nikolov, D. B., Chen, H., Halay, E., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G. & Burley, S. K. (1995) *Nature (London)* **377,** 119–128.