Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 13475–13480, December 1997
Biochemistry

Bidirectional binding of the TATA box binding protein to the TATA box

JULIA M. COX*, MATTHEW M. HAYWARD*, JENNIFER F. SANCHEZ*, LAURA D. GEGNAS*, SARINA VAN DER ZEE*, JONATHAN H. DENNIS†, PAUL B. SIGLER†, AND ALANNA SCHEPARTZ*‡

*Departments of Chemistry and †Molecular Biophysics and Biochemistry and the †Howard Hughes Medical Institute, Yale University, New Haven, CT 06520

Communicated by Peter B. Dervan, California Institute of Technology, Pasadena, CA, September 23, 1997 (received for review May 13, 1997)

ABSTRACT **By selective attachment of a DNA cleavage agent to specific residues in the yeast TATA box binding protein (yTBP), we demonstrate that, in solution, yTBP binds to the TATA boxes of both the adenovirus major late promoter and the yeast *CYC1* promoter with only a modest preference in orientation and binds well to several overlapping binding sites. The general factors TFIIA and TFIIB each increase the rotational and translational selectivity of yTBP but are not sufficient, at least individually, to confer a unique polarity to the preinitiation complex. We conclude that TBP alone cannot define the productive orientation of general factor assembly on a promoter.**

The TATA box binding protein (TBP) is required by all three RNA polymerases for the promoter-specific initiation of transcription (1). All eukaryotic TBP–DNA complexes observed in crystal structures show the conserved C-terminal domain of TBP (TBP_c) bound to the TATA box in a single orientation that is consistent with the assembly of a preinitiation complex with a unique polarity (2–8). The binding of TBP to the TATA box is thought to orient the complex correctly on the promoter and nucleate preinitiation complex formation (3, 9, 10). Several groups have proposed reasons why the highly symmetric TBP_c molecule binds to a nearly symmetric TATA box in only one of two possible orientations in the crystal. These reasons include amino acid and electrostatic differences between the C- and N-terminal repeats of TBP_c coupled with the differential deformability of each half of the TATA box (2, 3, 11, 12). Despite these factors that could favor the orientation of TBP_c observed in the crystal, the pseudo-2-fold symmetry of TBP_c and of many TATA boxes is intriguing. Approximately 80% of the amino acids that contact DNA are identical in the two halves of TBP_c, and molecular modeling studies reveal no unfavorable interactions when TBP is bound in the opposite orientation (13). Also, bidirectional transcription from a TATA box and forward transcription from a reverse TATA box can be observed, suggesting that perhaps TBP and/or the preinitiation complex can function bidirectionally (14–17). These issues led us to ask to what extent TBP binds to the TATA box in a unique orientation in solution.

The affinity cleavage method (18, 19) permits determination of the orientation of a protein bound to its DNA target site and the effects of other factors on this orientation (20). We used this technique to examine the orientation of TBP bound to a TATA box in solution. Using affinity cleavage, we discovered that TBP binds to the TATA boxes of the adenovirus major late promoter (AdMLP) and *CYC1* promoters with only a modest preference ($\Delta G_{obs} = 0.2 - 0.3$ kcal·mol^{-1}) in orientation. The general transcription factors TFIIB and TFIIA skew the ratio of TBP binding in favor of the generally accepted polarity but do not fix it to a unique orientation. The high degree of polarity achieved in regulated transcription (16, 17, 21) suggests that other promoter-specific factors define the polarity of the preinitiation complex.

MATERIALS AND METHODS

TBP Preparation and Derivitization. The TBP variants K97C and E188C were prepared as described by Kim *et al.* (3) and stored in 30 mM Tris (pH 7.5), 10% glycerol, 2 mM DTT, 50 mM KCl, and 1 mM EDTA. The fully functional deletion variant of yeast TFIIA used in crystallographic studies (yTFIIA′) was prepared as described (7). All TBP derivitization steps were performed at 4°C. Proteins (10–30 μM) were passed through a Sephadex G-25 column (Pharmacia) equilibrated with binding buffer (4 mM Tris, pH 8.0/5 mM $MgCl_2$/60 mM KCl/4% glycerol) and then added to a solution containing 10 equivalents of 5-iodoacetamido-1,10-phenanthroline (22), 3% dimethylformamide, 40 mM Tris (pH 8.0), 60 mM KCl, 4% glycerol, and 5 mM $MgCl_2$ and allowed to react for 15 h. Dimethylformamide and excess 5-iodoacetamido-1,10-phenanthroline were removed by gel filtration. Ellman's analysis (23) and carboxymethyl–cysteine amino acid analysis (24) indicated that one of the two cysteines on wild-type (WT) TBP and one of the three cysteines on K97C and E188C remained unalkylated after reaction. This cysteine was most likely C164, which is on a β-strand facing the interior of the protein. As long as specific cleavage is seen for K97C-1,10-phenanthroline (OP) and E188C-OP and not for WT-OP, the alkylation of a cysteine on WT TBP is inconsequential.

Footprinting and Cleavage Experiments. TBP (30 nM or 0.5–600 nM for cleavage and footprinting reactions, respectively) and 5′ end-labeled DNA (200 pM) were incubated for 30 minutes at 25°C with 0.04 mg/ml BSA and 0.12 mg/ml poly dG·dC in binding buffer containing 0.1% Nonidet P-40. For footprinting reactions, 0.03 units of DNase I was added to each binding reaction and quenched after 40 s by addition of 0.6 volumes formamide loading buffer. For cleavage reactions, a solution of 0.1% Nonidet P-40, 300 μM $CuSO_4$, 18 mM mercaptopropionic acid, and 18 mM H_2O_2 in 1/6 reaction volume of binding buffer was added and quenched after 3 h by ethanol precipitation. Radioactivity was quantified using a Betascope 603 Blot Analyzer (Betagen, Waltham, MA) or a Storm 840 PhosphorImager (Molecular Dynamics). Reactions containing $yTFIIB_c$ were carried out as above except $yTFIIB_c$ was added in 20 mM Hepes (pH 7.9), 1 mM EDTA, 10% glycerol, 1 mM DTT, 1 mM

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: TBP, TATA box binding protein; TBP_c, conserved C-terminal domain of TBP; AdMLP, adenovirus major late promoter; WT, wild-type.

‡To whom reprint requests should be addressed. e-mail: alanna@milan.chem.yale.edu.

phenylmethylsulfonyl fluoride, 0.15 M potassium acetate, 50 mg/ml BSA, and 0.1% Nonidet P-40. Binding reactions containing yTFIIA′ were carried out as above, except TFIIA′ was added in 30 mM Tris (pH 8.0), 10% glycerol, 5 mM DTT, 300 mM KCl, and 2 mM EDTA. Cleavage ratios were calculated by summing the cleavage intensities on either side of the center of the 8-bp TATA box and were corrected for background cleavage.

RESULTS AND DISCUSSION

Two yeast TBP variants, K97C and E188C, were prepared to examine the orientational specificity of the TBP–TATA box interaction. Each variant was modified chemically to carry an OP DNA-cleaving agent at a position on either the N- or the C-terminal stirrup, respectively (Fig. 1*A*). The resulting molecules, K97C-OP and E188C-OP, bound DNA containing the AdMLP TATA box to form complexes with dissociation constants of 8 ± 4 nM and 2 ± 2 nM, respectively, under the conditions described. These values are comparable to the dissociation constant of the WT yTBP·AdMLP complex (2 ± 2 nM), indicating that these alkylated TBP derivatives are appropriate models for WT yTBP binding in solution. The rationale behind our experiments (Fig. 1*B*) is as follows: If K97C-OP binds the TATA box in solution in the single orientation predicted by crystallography, then it should cleave DNA on only the downstream side of the TATA box; similarly, if E188C-OP binds the TATA box in the single orientation predicted by crystallography, then it should cleave DNA on only the upstream side of the TATA box. If either alkylated protein binds DNA in two orientations, however, then cleavage on both sides of the TATA box should be observed (20).

The alkylated proteins K97C-OP and E188C-OP were bound individually to an 80-bp DNA fragment containing the AdMLP TATA box after which cupric sulfate and a reducing agent were added to initiate DNA cleavage. Each protein cleaved the DNA on both sides of the TATA box with a 60–67% preference for the side expected based on data from crystallography (2, 3) (Fig. 2 *A* and *B*). Similar cleavage ratios were observed when K97C-OP and E188C-OP were used to cleave DNA containing the *CYC1* TATA box (Fig. 2*C*). The shapes of the affinity cleavage patterns produced on these

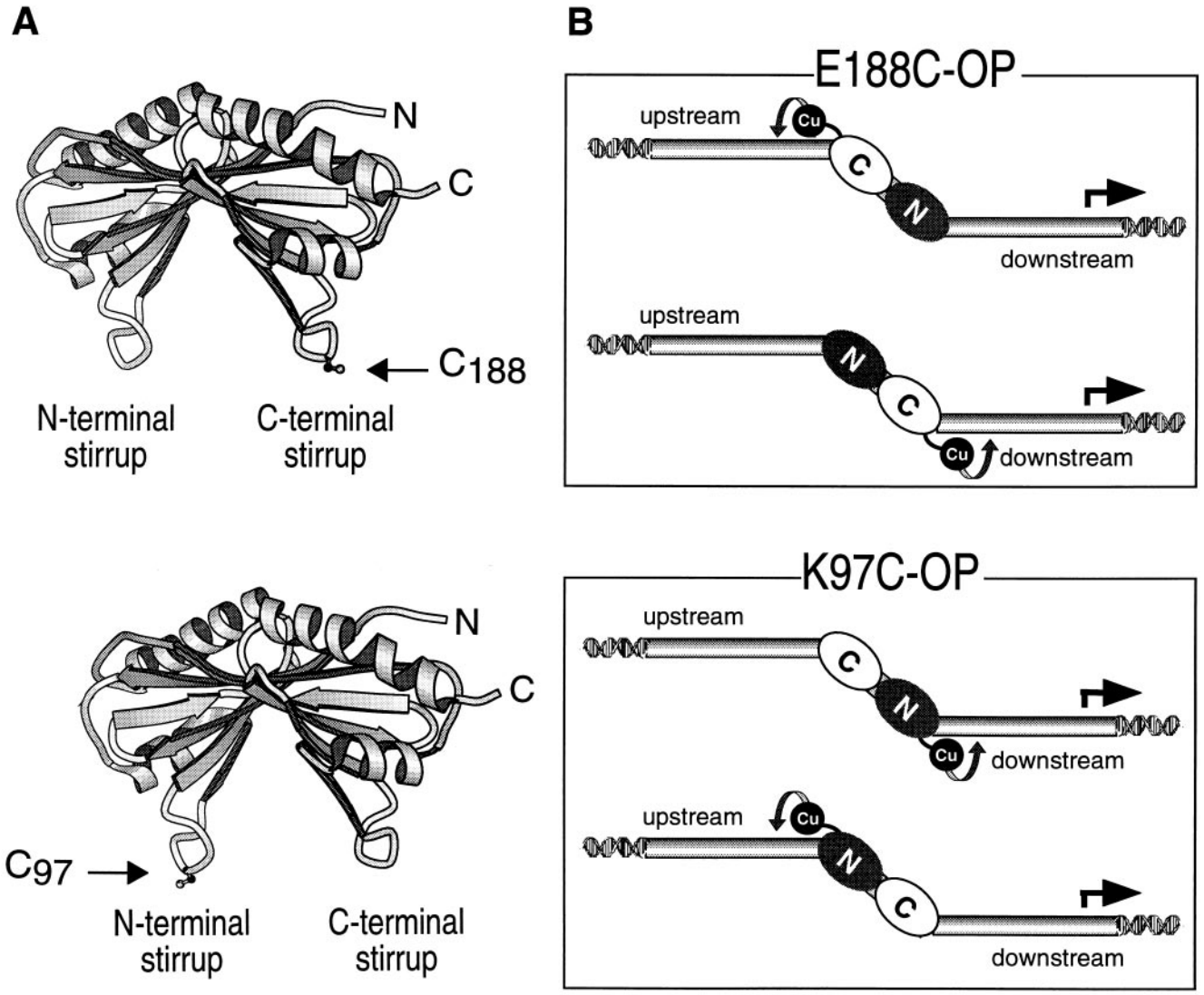

FIG. 1. Strategy for affinity cleavage of DNA by TBP K97C-OP and E188C-OP. (*A*) K97C and E188C yTBP. These two variants were chosen because they allow a DNA cleavage agent to be placed on equivalent positions of the N- and C-terminal stirrups, respectively, with a minimum of steric interference. (*B*) Experimental strategy. Scheme illustrating the regions of the promoter that are expected to be cleaved in the specified orientation. N and C refer to the pseudosymmetrical halves of TBP_c formed by the N- and C-terminal repeats, respectively. All crystal structures of eukaryotic TBP_c-containing promoter complexes show the N-terminal domain of TBP_c facing downstream toward the start site of transcription. The lengths of the flanking DNA are not drawn to scale.

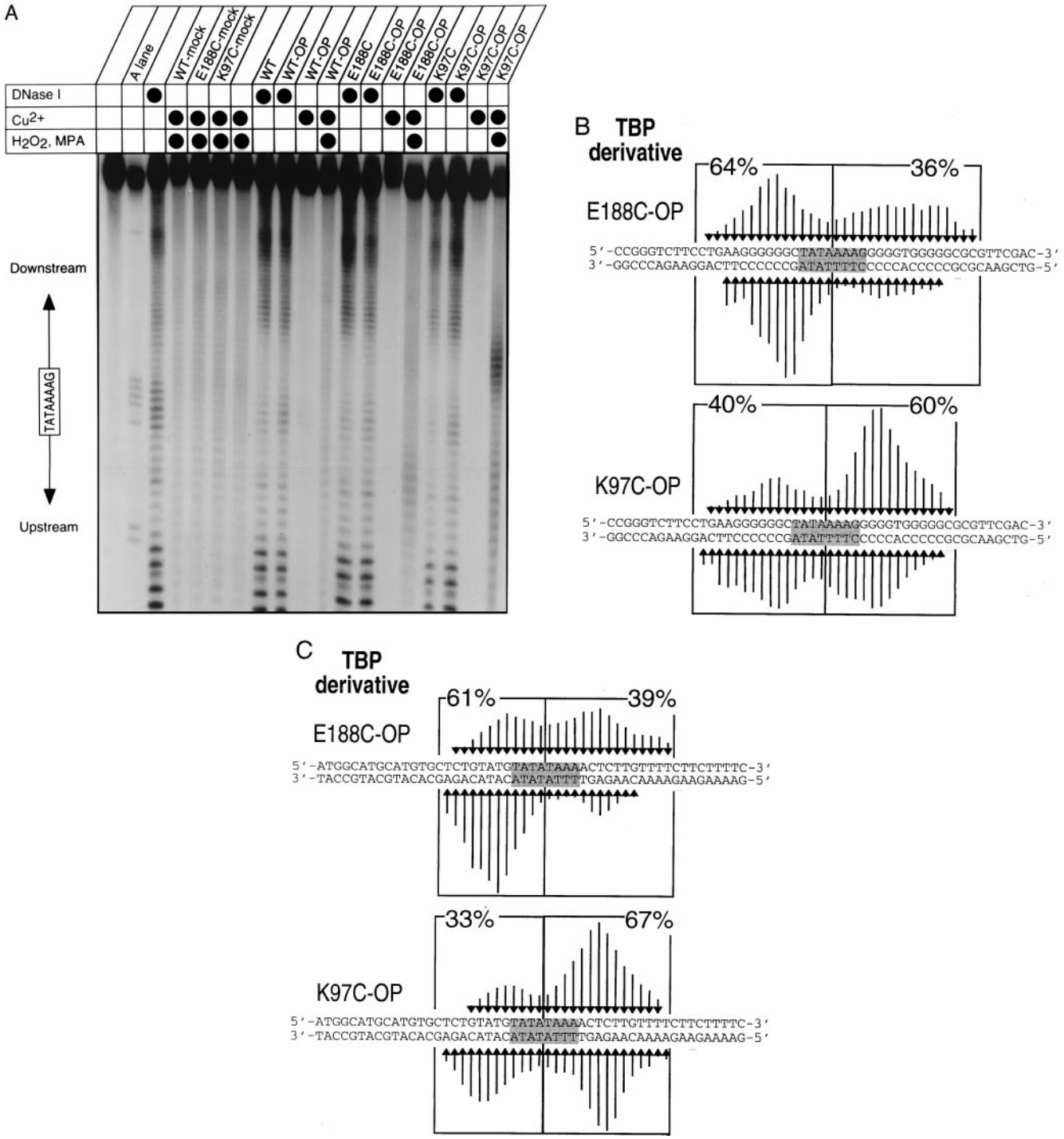

FIG. 2. Affinity cleavage (18, 19) of the AdMLP TATA box by yTBP derivatives. (*A*) Autoradiogram of a 12% high resolution sequencing gel analyzing the products of affinity cleavage of the coding strand of an 80-bp fragment of the AdMLP. The diagram on the left shows the sequence of the AdMLP TATA box; "upstream" and "downstream" refer to the location of the DNA relative to the TATA box. The left-most lane shows a DNA control. Histograms illustrating cleavage by E188C-OP and K97C-OP at each base pair in the AdMLP (*B*) and *CYC1* (*C*) promoters. The extent of cleavage at each position (above a cupric ion, hydrogen peroxide, and mercaptopropionic acid control) is proportional to the length of the arrow. The TATA box is shaded gray.

TATA boxes were similar, suggesting that both binary complexes possessed equivalent architectures. K97C-OP and E188C-OP cleaved DNA equally on either side of a symmetric TATA box located within the context of the AdMLP sequence (data not shown). This result suggests that the small degree of preferential orientation observed with the AdMLP and *CYC1* TATA boxes is encoded within the 8-bp TATA box sequence, whereas the influence of the flanking DNA sequences on TBP binding *per se* is negligible.

Several control experiments were performed to establish that the observed cleavage patterns resulted from a site-specifically alkylated yTBP derivative. No specific DNA cleavage was observed using E188C and K97C variants that had been treated in mock alkylation reactions with *o*-phenanthroline, a reagent that cannot alkylate cysteine residues (Fig. 2*A*). No DNA cleavage was observed when the yTBP derivative, cupric sulfate, or reducing agent was omitted from the cleavage reaction, and no specific DNA cleav-

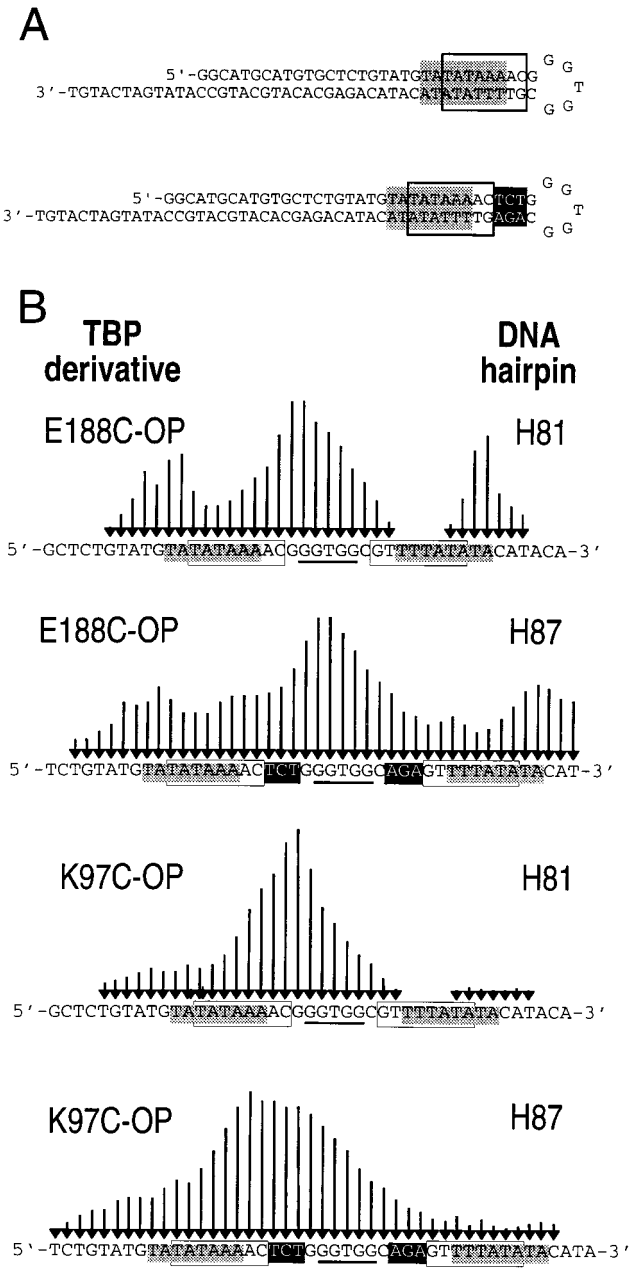

FIG. 3. Affinity cleavage patterns indicating the axial imprecision of yTBP binding. (*A*) Sequences of H81 (*Upper*) and H87 (*Lower*) shown in hairpin form. There are two overlapping TATA boxes in the *CYC1* sequence displaced by 2 bp; here, one is shaded gray and the other is outlined. The extra base pairs added between the TATA region and the hairpin loop are highlighted in black on H87. (*B*) Histograms illustrating cleavage of H81 and H87 by E188C-OP and K97C-OP. Quantitation was performed as described in Fig. 2 *B* and *C*. The DNA is shaded as in *A*, and the hairpin loop is underlined.

age was observed with WT-OP under any conditions (Fig. 2*A*). The extent of DNA cleavage on the upstream and downstream sides of the TATA box increased at equal rates over time in the presence of both K97C-OP and E188C-OP (data not shown). Finally, to insure that our DNA cleavage experiments reported the geometry of a single, specific yTBP·DNA complex, we performed the cleavage reaction, then partitioned the free and yTBP-OP-bound forms of the AdMLP DNA on a denaturing gel (25). When the DNA in the bound bands were run on a denaturing gel, the cleavage patterns observed were identical to those seen in standard experiments. This experiment indicated that cleavage was due to a 1:1 yTBP·DNA complex and that gel electrophoresis did not alter the ratios of yTBP·DNA isomers present in solution (data not shown). Taken with the data shown in Fig. 2, these experiments demonstrate that, in the absence of other protein interactions, yTBP cannot orient itself uniquely on the AdMLP or *CYC1* TATA box. Under these conditions, yTBP binds DNA as a mixture of orientational isomers that are related by a 180° rotation about the pseudo dyad of the complex.

The cleavage patterns generated by E188C-OP and K97C-OP each span two turns of the DNA helix and as such are considerably broader than those observed for most protein–OP conjugates (26). Because DNA cleavage by tethered OP derivatives is mediated by a nondiffusible copper-oxo species, the unusually broad patterns cannot result from cleavage by a diffusible cleavage agent (27). Two explanations exist for the spread in the cleavage patterns. The first is that flexibility within the derivatized DNA-bound yTBP molecule increases the number of deoxyribose residues within reach of the tethered phenanthroline complex. A second explanation is that yTBP binds to a small continuum of overlapping binding sites centered on the 8-bp TATA box. To discriminate between these possibilities, we designed a DNA hairpin (H81) that restricted the axial mobility of bound yTBP (Fig. 3*A*). If the broad cleavage patterns result from translation of yTBP along the DNA helix axis, then, for reasons pointed out by Geiger *et al.* (7), we would expect cleavage of H81, in which the TATA box abuts the hairpin loop, to produce a narrower pattern than cleavage of H87, in which the TATA box is separated from the loop by 3 bp. On the other hand, if the broad cleavage pattern is the result of conformational flexibility within the bound yTBP derivative, then the cleavage patterns on H81 and H87 should be comparable. Cleavage of H87 by either K97C-OP or E188C-OP produced a significantly broader pattern than did cleavage of H81 (Fig. 3*B*). These data indicate that, in the absence of other protein factors, yTBP does not position itself precisely on the TATA box. The cleavage observed on either side of the TATA box cannot result from a protein bound in a unique orientation but translated axially along the strand, however, as the center of the TATA box is protected from affinity cleavage and a clear DNase I footprint is observed within the TATA box. Note that the cleavage patterns generated by the yTBP derivatives on the *CYC1* and AdMLP TATA boxes were virtually identical (Fig. 2) despite the fact that the *CYC1* promoter contains two overlapping 8-bp TATA boxes (Fig. 3*A*) and the AdMLP does not. These results emphasize that the thermodynamic boundaries of the TATA box are not well defined. The TBP·TATA box interface is unique because axial displacement of TBP by 1 or 2 bp does not place the TBP binding surface significantly out of register with its complementary DNA target sequence (28).

The observation that yTBP binds DNA as a mixture of orientational and axial isomers led us to question whether specificity might be provided by other components of the preinitiation complex. Crystallographic and biochemical experiments show that TFIIB and TFIIA both interact with the TBP_c promoter complex (5, 7, 8, 29–31). By contacting both the DNA and protein members of this partnership, TFIIB and TFIIA could help orient TBP in one direction and/or position TBP axially on a single TATA box. To evaluate the roles of these general factors in the orientation of yTBP binding, we performed cleavage experiments in the presence of the C-terminal protease-resistant core of yeast TFIIB ($yTFIIB_c$) or the fully functional deletion variant of yeast TFIIA (yTFIIA′) used in crystallographic studies (7). TBP derivatives were chosen to avoid steric interference with TFIIA (E188C-OP) or TFIIB (K97C-OP). Cleavage experiments were performed under conditions in which an ex-

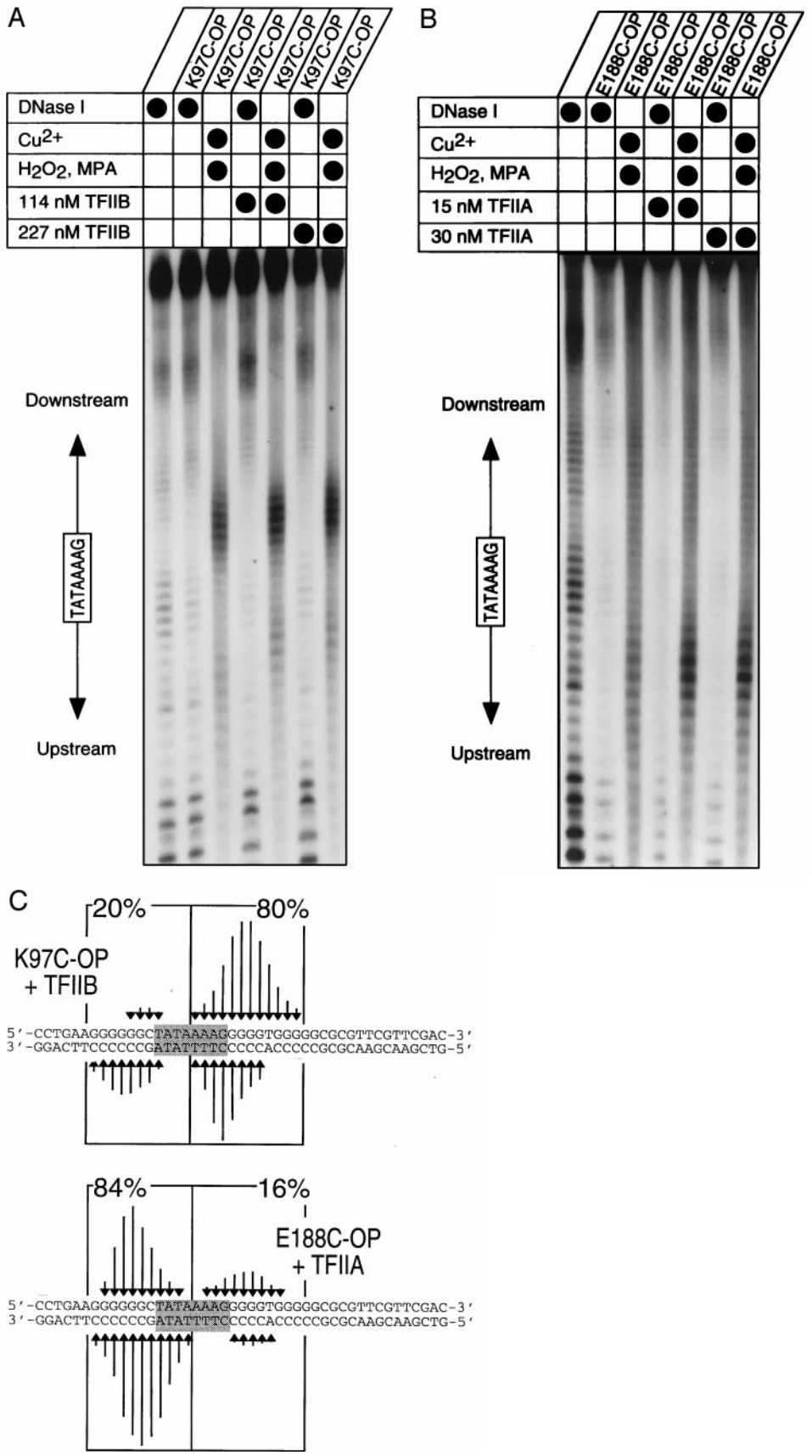

FIG. 4. Affinity cleavage of DNA by yTBP derivatives in the presence of $yTFIIB_c$ or yTFIIA′. (*A*) Autoradiogram of a 12% high resolution sequencing gel analyzing the products of affinity cleavage of the top strand of an 80-bp fragment of the AdMLP by K97C-OP in the presence of $yTFIIB_c$. (*B*) Autoradiogram of a 12% high resolution sequencing gel analyzing the products of affinity cleavage of the top strand of an 80-bp fragment of the AdMLP by E188C-OP in the presence of yTFIIA′. (*C*) Histograms illustrating cleavage by K97C-OP at each base pair surrounding the AdMLP box in the presence of $yTFIIB_c$ or by E188C-OP at each base pair surrounding the AdMLP box in the presence of yTFIIA′.

panded DNase I footprint and enhanced binding of yTBP to DNA established that yTFIIB$_c$ or yTFIIA′ was bound (Fig. 4 *A* and *B*). Incubation of the ternary complex containing K97C-OP, yTFIIB$_c$, and the AdMLP with cupric sulfate and reducing agent cleaved DNA on both sides of the TATA box but reflected a significantly greater preference for the ''correct'' orientation than that seen for TBP alone (80 ± 3% vs. 60 ± 4%) (Fig. 4*A*). In addition, the distribution of cleavage sites in the presence of yTFIIB$_c$ was narrower than in its absence, indicating that yTFIIB$_c$ enhanced the specificity of yTBP for the 8-bp TATA box (Fig. 4 *A* and *C*). Similarly, yTFIIA′ shifted the orientational preference of TBP toward the correct side (84 ± 4% vs. 64 ± 4%) and narrowed the E188C-OP cleavage pattern (Fig. 4 *B* and *C*). Thus, although yTFIIB$_c$ and yTFIIA′ helped orient yTBP correctly on the TATA box, 20% and 16% of the yTBP·yTFIIB$_c$·TATA box and yTBP·yTFIIA′·TATA box complexes, respectively, still bound in a inverted orientation. It appears that TFIIA′ and TFIIB$_c$ enhanced the orientational preference and axial positioning of TBP on the promoter by exploiting heretofore unrecognized sequence-specific differences in the TATA box and/or the flanking DNA or by enhancing the intrinsic selectivity of TBP itself. Further structural and biochemical studies will be required to fully understand these effects.

Several factors have been proposed to explain the orientational specificity of TBP$_c$ observed by crystallography (2, 3, 11, 12). These factors include amino acid and electrostatic differences between the C- and N-terminal repeats of TBP$_c$ coupled with the differential deformability of the DNA in each half of the TATA box. These factors may explain why the correct orientation is favored, but the preference for binding in this orientation is small, reflecting a difference in free energy of only 0.2–0.3 kcal·mol^{-1}. Our finding that yTBP binds the TATA box in two orientations is consistent with simulations that reveal no unfavorable steric interactions in the interface of the ''incorrect'' TBP$_c$·TATA box complex (13). Our results also are consistent with some transcription experiments that demonstrate bi-directional transcription from an asymmetric TATA box in a pol III promoter and transcription from a reverse TATA box in a pol II promoter (14, 15). A crystal lattice demands a uniform complex, and it appears that, in the TBP-containing structures determined to date (2, 3, 5–8), the lattice prefers the marginally more stable TBP·TATA box complex. Although the kinetics of TBP binding may be slow on a physiological time scale, they may be fast enough for the equilibrium in solution to be drawn progressively to a single orientation by transfer to a stable, insoluble crystal.

In summary, these results demonstrate that the asymmetry present in TBP$_c$ and most TATA boxes is not sufficient to determine the directionality of preinitiation complex formation. Although the two general factors we examined (TFIIB and TFIIA) stabilized the preferred orientation, they did not, at least individually, completely fix the orientation of yTBP. It is likely that productive orientation of the transcriptional complex is aided further by other components of the preinitiation complex (or the holoenzyme itself), such as activators, co-activators, and accessory factors. It is not clear what level of incorrect polarity can be tolerated in the assembly of the preinitiation complex at any particular promoter. To the extent that correct polarity is required, however, it is likely to be achieved in increments of favorable binding free energy provided by accretion of components that assemble in the transcriptionally productive direction.

We are grateful to David Sigman for a generous sample of 5-iodoacetamido-1,10-phenanthroline and to Steve Hahn for yTFIIB$_c$. This work was supported by grants from the National Institutes of Health to P.B.S. (GM15225) and A.S. (GM52544 and GM53829).

1. Hernandez, N. (1993) *Genes Dev.* **7,** 1291–1308.
2. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993) *Nature (London)* **365,** 520–527.
3. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993) *Nature-(London)* **365,** 512–520.
4. Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N. H., Roeder, R. G. & Burley, S. K. (1992) *Nature (London)* **360,** 40–46.
5. Nikolov, D. B., Chen, H., Halay, E. D., Asheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G. & Burley, S. K. (1995) *Nature (London)* **377,** 119–128.
6. Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G. & Burley, S. K. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4862–4687.
7. Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. (1996) *Science* **272,** 830–836.
8. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. (1996) *Nature(London)* **381,** 127–134.
9. Chatterjee, S. & Struhl, K. (1995) *Nature (London)* **374,** 820–822.
10. Klages, N. & Strubin, M. (1995) *Nature (London)* **374,** 822–823.
11. Suzuki, M., Allen, M. D., Yagi, N. & Finch, J. T. (1996) *Nucleic Acids Res.* **24,** 2767–2773.
12. Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996) *J. Mol. Biol.* **261,** 239–254.
13. Miaskiewicz, K. & Ornstein, R. L. (1996) *J. Biomol. Struct. Dyn.* **13,** 593–600.
14. Whitehall, S. K., Kassavetis, G. A. & Geiduschek, E. P. (1995) *Genes Dev.* **9,** 2974–2985.
15. Li, J. J., Kim, R. H. & Sodek, J. (1995) *Biochem. J.* **310,** 33–40.
16. Xu, L. C., Thali, M. & Schaffner, W. (1991) *Nucleic Acids Res.* **19,** 6699–6704.
17. Huang, W., Wong, J. M. & Bateman, E. (1996) *Nucleic Acids Res.* **24,** 1158–1163.
18. Dervan, P. B. (1986) *Science* **232,** 464–471.
19. Sigman, D. S., Bruice, T. W., Mazumder, A. & Sutton, C. L. (1993) *Acc. Chem. Res.* **26,** 98–104.
20. Chen, L., Oakley, M. G., Glover, J. N. M., Jain, J., Dervan, P. B., Hogan, P. G., Rao, A. & Verdine, G. L. (1995) *Curr. Biol.* **5,** 882–889.
21. O'Shea-Greenfield, A. & Smale, S. T. (1992) *J. Biol. Chem.* **267,** 1391–1402.
22. Chen, C.-H. B. & Sigman, D. S. (1987) *Science* **237,** 1197–1201.
23. Ellman, G. L. (1959) *Arch. Biochem. Biophys.* **82,** 70–77.
24. Moore, S. & Stein, W. H. (1963) *Methods Enzymol.* **6,** 819–831.
25. Sutton, C. L., Mazumder, A., Chen, C. B. & Sigman, D. S. (1993) *Biochemistry* **32,** 4225–4230.
26. Pan, C. Q., Landgraf, R. & Sigman, D. S. (1994) *Mol. Microbiol.* **12,** 335–343.
27. Gallagher, J., Chen, C. H., Pan, C. Q., Perrin, D. M., Cho, Y. M. & Sigman, D. S. (1996) *Bioconjugate Chem.* **7,** 413–420.
28. Coleman, R. A. & Pugh, B. F. (1995) *J. Biol. Chem.* **270,** 13850–13859.
29. Lee, S. & Hahn, S. (1995) *Nature (London)* **376,** 609–612.
30. Imbalzano, A. N., Zaret, K. S. & Kingston, R. E. (1994) *J. Biol. Chem.* **269,** 8280–8286.
31. Lagrange, T., Kim, T. K., Orphanides, G., Ebright, Y. W., Ebright, R. H. & Reinberg, D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 10620–10625.