1966–1973 Nucleic Acids Research, 1994, Vol. 22, No. 11

The upstream activator CTF/NF1 and RNA polymerase II share a common element involved in transcriptional activation

Hua Xiao[1,2,+], John T.Lis[2,*], Hua Xiao[3], Jack Greenblatt[3] and James D.Friesen[1]
Department of Genetics, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada, [2]Section of Biochemistry, Molecular and Cell Biology, Biotechnology Building, Cornell University, Ithaca, NY 14853, USA and [3]Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, Ontario M5G 1L6, Canada

Received April 14, 1994; Accepted April 21, 1994

ABSTRACT

The carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II consists of tandem repeats of a heptapeptide with the consensus YSPTSPS. It has been shown that the heptapeptide repeat interacts directly with the general transcription factor TFIID. We report here that the CTD activates transcription when fused to the DNA-binding domain of GAL4. More importantly, we find that the proline-rich transcriptional activation domain of the CCAAT-box-binding factor CTF/NF1 contains a sequence with striking similarity to the heptapeptide repeats of the CTD. We show that this CTD-like motif is essential for the transcriptional activator function of the proline-rich domain of CTF/NF1. Deletion of and point mutations in this CTD-like motif abolish the transcriptional activator function of the proline-rich domain, while natural CTD repeats from RNA polymerase II are fully functional in place of the CTD-like motif. We further show that the proline-rich activation domain of CTF/NF1 interacts directly with the TATA-box-binding protein (TBP), and that a mutation in the CTD-like motif that abolishes transcriptional activation reduces the affinity of the proline-rich domain for TBP. These results demonstrate that a class of proline-rich activator proteins and RNA polymerase II possess a common structural and functional component which can interact with the same target in the general transcription machinery. We discuss the implications of these results for the mechanisms of transcriptional activation in eucaryotes.

INTRODUCTION

Transcription by RNA polymerase II is a multi-step process that begins with an ordered assembly of general transcription factors and the RNA polymerase on a promoter (1). For genes with a TATA-box, the first step is the binding of the general transcription factor TFIID to the TATA-box, possibly with the aid of TFIIA, followed by binding of TFIIB to form a promoter/DAB complex (1). RNA polymerase II, with its associated proteins such as RAP30/74 or TFIIF (2) and other factors, is then able to associate with this promoter complex to form a transcription-competent complex, which is capable of a basal level of transcription (reviewed in 3, 4). This basal transcription can be modulated either positively or negatively by regulatory proteins that bind to sites near or distant from the basal transcription machinery. Activator proteins that greatly increase the rate of transcription usually consist of two separable, functionally autonomous domains (5, 6). One domain is involved in recognition of and binding to specific DNA sequence elements near or distant from the promoter. The other is required for activation of transcription. However, the mechanisms by which activator proteins stimulate transcription are not clear. Presumably, any one of the steps leading to productive transcription could be rate-limiting and thus a potential target for the action of activator proteins. Ptashne and colleagues have proposed that the activation domains of acidic activator proteins stimulate transcription by making direct contact with a component(s) of the general transcription machinery, for example, one or more of the general transcription factors (5–7). Recent studies have demonstrated that *in vitro* an acidic activation domain is indeed capable of binding to two components of the general transcription machinery, TFIID and TFIIB, and the affinity of the acidic activation domain for the general transcription factor TFIID or TFIIB correlates with its ability to activate transcription (8–11). It thus appears that a critical feature of a transcriptional activation domain is its ability to make direct contact with a component(s) of the general transcription machinery.

The carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II (12, 13) has been suggested to play a role in the communication between upstream activator proteins and RNA polymerase II. Previous studies have shown that

*To whom correspondence should be addressed

+Present address: Section of Biochemistry, Molecular and Cell Biology, Biotechnology Building, Cornell University, Ithaca, NY 14853, USA

transcription by RNA polymerase II from different promoters requires different numbers of the heptapeptide repeats in the CTD (14), and that the requirements for the number of heptapeptide repeats also vary, depending on the activity of a particular upstream activator being used at a promoter (15). More recently, the CTD has been shown to interact directly with the TATA-binding subunit (TBP) of the general transcription factor TFIID (16, 17). In this study, we attempt to determine the functional relationship between upstream transcriptional activators and the CTD of RNA polymerase II by testing the ability of the CTD to stimulate transcription when fused to a DNA-binding domain. Our results show that the CTD of RNA polymerase II can function upstream as an activation domain for transcription when fused to the DNA binding domain of GAL4. More importantly, the results have led to the unexpected finding that CTD-like repeats are an essential functional component of the proline-rich activation domain of a natural upstream activator, CTF/NF1.

MATERIALS AND METHODS

Media and strains

Media were prepared according to standard methods (18). The medium used for galactose induction is synthetic medium with 2% galactose, 2% ethanol, and 2% glycerol. Strain GGY1:171 (19) used for assays of GAL4 derivatives was obtained from Mark Ptashne's laboratory. This strain lacks the entire GAL4 coding sequence.

Site-directed mutagenesis and plasmid constructions

All DNA manipulations were carried out by using standard methods (20). Alterations of the CTD of RNA polymerase II of the yeast *Saccharomyces cerevisiae* were made first in a KpnI/EcoRI fragment of the yeast *RPO21* gene, which contains the entire CTD coding region and the transcription termination site (21). This DNA fragment was subcloned into KpnI and EcoRI digested plasmid vector pTZ19R to generate pYCD. For the purpose of subsequent subcloning, an XhoI restriction site was introduced into the SpeI site which is downstream of the *RPO21* transcription-termination site to generate pYCDx. An NruI site was created in the eighth repeat of the CTD in pYCDx to generate pYCDR8nr by changing a C to a G at position 5078 of RPO21 by site-directed mutagenesis; a SalI site was created at the 3′ side of the final repeat of the CTD by changing the CAA to GTC at positions 5452 to 5454 to generate pYCDsh; and a BamHI site was created in the twenty-fourth repeat by changing the ATC to GGA at positions 5409 to 5411 to generate pYCD24b.

The GAL4 derivatives were constructed in the yeast vector pMA424 (22), or pHGX1 that contains GAL4 residues 1 to 147 under the control of the yeast heat shock factor promoter. To construct pHGX1, pMA424 was digested with restriction enzymes HindIII (14 bp upstream of the ATG codon of GAL4) and SalI (following codon 147 of GAL4). This HindIII/SalI fragment was fused to an EcoRI/PvuII fragment from the 5′ upstream region of the yeast heat shock factor gene (23). This fusion DNA was then cloned into the BamHI (thus destroyed) and SalI sites of plasmid pRS313 (24).

Plasmid pGYR26 was constructed by inserting a BsiWI (filled in with Klenow)/XhoI fragment encoding the entire yeast CTD from pYCDR8nr into the EcoRI (filled in with Klenow) and SalI sites of pMA424 (Note that the restriction enzyme BsiWI cuts at nineteen codons before the consensus CTD repeats). pGYR8 was constructed by inserting the oligonucleoide CGA*TGA*-TATAG with an in frame stop codon (in italics) into the NruI site of pYCDR8nr and subsequent subcloning of the BsiWI (filled in with Klenow)/XhoI fragment containing the stop codon into EcoRI (filled in with Klenow) and SalI digested pMA424. Plasmids pGDR36 and pGDR18 were constructed as follows. A SalI/EcoRI (filled in with Klenow) fragment encoding the heptapeptide repeats seven through forty-two of the *Drosophila* CTD (25) was first inserted between the NruI and SnaBI sites of pYCDR8nr to generate plasmid pYCDR8dR36. This cloning recreated the SalI site and placed the transcription termination site of yeast RPO21 downstream of the *Drosophila* CTD coding sequence. A SalI (filled in with Klenow)/XhoI fragment from pYCDR8dR36 was cloned into BamHI (filled in with Klenow) and SalI digested pMA424 to generate pGDR36. An EcoRI/ScaI fragment encoding eighteen *Drosophila* CTD repeats from pGDR36 and a SalI (filled in with Klenow)/XhoI fragment from pYCDsh were inserted into the EcoRI and SalI sites of pMA424 to generate pGDR18.

Plasmid pGF399-499 (see Figure 2) was constructed by inserting a BglII (filled in with Klenow)/HindIII fragment that encodes the proline-rich activation domain of CTF/NF1 (26, 27) and a HindIII/XhoI fragment containing the transcription termination site of the yeast HSF into BamHI (filled in with Klenow) and SalI digested pHGX1. Plasmids pGF399-438 was constructed as follows. A KpnI site with the sequence CCGGGTACCCGG was first fused to the HincII site of pYCDsh and then digested with KpnI. After removal of the KpnI protruding end with T4 DNA polymerase, the plasmid was digested with XhoI. A fusion of this fragment and an EcoRI/SmaI fragment encoding amino acids 399 to 438 of CTF/NF1 from pGF399-499 was cloned into the EcoRI and SalI sites of pHGX1 to generate pGF399-438 in which the final twelve codons of yeast *RPO21* are in frame with *CTF/NF1*. To construct pGF399-438YR1, pYCD24b was first digested with BamHI (which cuts once in the twenty-fourth repeat of the CTD and once in the polylinker of the vector) and then religated to generate pYCDR1. pYCDR1 was then digested with HincII and XhoI to generate a fragment encoding the final one and 5/7 CTD repeats of the yeast *RPO21*. This fragment and an EcoRI/SmaI fragment encoding amino acids 399 to 438 of CTF/NF1 from pGF399-499 were cloned into EcoRI and SalI digested pHGX1 to generate pGF399-438YR1. pGF462PR was constructed by changing codon 462 from CCG (proline) to CGA (arginine) by site-directed mutagenesis, which also created an NruI site between codons 461 and 462 of CTF/NF1. pGFΔ462-485 was constructed by deleting codons 462 to 485 between the NruI and BamHI sites of pGF462PR. The following derivatives contain substitutions between codons 461 (NruI site) and 486 (BamHI site) of CTF/NF1: pGFΔ462-485yR3 contains a sequence encoding heptapeptide repeats 9 to 11 of the yeast CTD, pGFΔ462-485yR8 contains a sequence encoding heptapeptide repeats 1 to 8 of the yeast CTD, and pGFΔ462-485TS contains a sequence encoding 21 amino acids, RTTTTITITTTSGSWEGFQNS.

Fusions of the wild type and mutant derivatives of the proline-rich domain of CTF/NF1 to the maltose-binding protein were made by subcloning an EcoRI/PstI fragment containing the proline-rich domain or the mutant derivatives into EcoRI and PstI digested pMAL-c2 (New England Biolabs, Inc.).

To construct a GST-yeast TBP fusion, an EaeI(cuts at the second codon of yTBP)/BamHI fragment from plasmid pRS314-R12D (obtained from Dr Tony Weil at Vanderbilt University) was inserted into NaeI/BamHI digested plasmid

pGdTBP in which the first 45 codons of dTBP was fused to the second codon of yTBP.

Yeast transformation and β-galactosidase assays

Plasmid constructs were transformed into yeast as described (28). Transformed yeast cells were grown in selective medium to an A600 of 1.0−1.5. Cells were centrifuged in a 1.5ml Eppendorf tube in a microfuge for 2 minutes. The cell pellets were resuspended in β-galactosidase assay buffer (50mM KPO_4, pH7.4, 1mM $MgCl_2$), and cells were permeabilized with chloroform and sodium dodecyl sulfate as previously described (29). β-galactosidase activity was determined by using the substrate Chlorophenol red-β-D-galactopyranoside (Boehringer Mannheim) as described (30).

Protein blot analysis of yeast whole cell extracts

Yeast whole cell extracts were prepared as described (14). The extracts were further treated with lambda phosphatase, and analyzed by protein blotting using antibody against the GAL4 DNA-binding domain (a gift from Dr Mark Ptashne's laboratory) and the ECL detection system from Amersham.

Protein affinity chromatography

The proline-rich domain (amino acids 399−499) of CTF/NF1 and its derivative CTFD462-485TS (see Plasmid construction for details) were fused to the *E.coli* maltose-binding protein in the expression vector pMAL-c2 (New England Biolabs, Inc.). The plasmids were transformed into *E.coli* strain BL21 (Novagen). The fusion proteins were induced and partially purified according to a protocol provided by the manufacturer (New England Biolabs, Inc.). Before use, the partially purified proteins were dialyzed extensively in ACB, the affinity-chromotography buffer (20mM HEPES, pH7.5, 5mM $MgCl_2$, 0.5mM EDTA, 0.2mM DTT, 0.2mM PMSF, 10% glycerol) containing 50mM KCl.

BL21 cells transformed with GST or GST/yTBP fusion plasmid were grown in YT to an OD600 of 0.7, and then induced with 0.2mM IPTG for 5 hours at 37°C. Cells were centrifuged, pellets resuspended in buffer A (20mM Tris.Cl, pH7.5, 1M NaCl, 1mM EDTA, 1mM DTT, 0.5mM PMSF), and frozen at −20°C overnight. Cells were then thawed and broken by sonication. Following contrifugation at 10,000g for 30 min., the supernatant was loaded onto a 0.6ml GST column previously equilibrated with buffer A. After extensive washes with buffer A, the columns were then equilibrated with ACB containing 50mM KCl. The concentration of GST or GST/yTBP on the column was later determined to be about 4mg/ml of resin. 1.5ml of the partially purified MBP/CTF1 fusion proteins at 1.0−2.0mg/ml was loaded onto the GST or GST/yTBP columns at a flow rate of about 1.5ml/hour. The columns were washed with 5 to 10 column volumes of ACB containing 50mM KCl, and bound proteins were eluted with ACB/50mM KCl/10mM GSH. The eluted proteins were detected by protein blotting using an anti-MBP antibody (purchased from New England Biolabs, Inc.).

Synthesis of a CTD peptide and competition analysis of CTF1 interaction with TBP

14-mer SPTSPSYSPTSPSY containing two consensus heptapeptide repeats of the CTD was synthesized at Cornell University Biotechnology Facility. The purity of the peptide was determined to be greater than 95%. For competition analysis of CTF1 interaction with TBP, the MBP-CTF1 fusion protein prebound to a GST-yeast TBP column was eluted either with ACB/50mM KCl/1.5mg ml^{-} of the synthetic CTD peptide, or with the egg white lysozyme (Boehringer Mannheim) as a control. The eluted proteins were detected by protein blotting using an anti-MBP antibody (purchased from New England Biolabs, Inc.).

RESULTS

The CTD of RNA polymerase II can act upstream as an activation domain for transcription

We constructed hybrid proteins that contain the CTD from the yeast *Saccharomyces cerevisiae* RNA polymerase II fused to the DNA-binding domain of the activator protein GAL4, and tested the ability of the hybrid proteins to stimulate transcription in yeast. As shown in Figure 1, the GAL4−CTD fusion proteins strongly stimulated transcription from a *gal1−lacZ* reporter gene. A hybrid protein containing the entire yeast CTD with twenty-six heptapeptide repeats stimulated transcription from the *gal1−lacZ* reporter gene by more than 100-fold (Figure 1 *b*). A hybrid protein containing eight of the twenty-six heptapeptide repeats stimulated transcription by about 10-fold (Figure 1 *c*). Thus, the level of transcriptional activation correlated roughly with the length of the CTD. Interestingly, the activity of RNA polymerase II on many promoters also correlates with the length of its CTD (14, 15, 31−35).

Since the CTD is a conserved domain of RNA polymerase II in eukaryotes, we used the same assay to test the ability of the CTD of *Drosophila melanogaster* RNA polymerase II to stimulate transcription in yeast. As shown in Figure 1 *d* and *e*, the hybrid protein pGDR36, which contains the C-terminal thirty-six heptapeptide repeats of the *Drosophila* CTD, stimulated transcription more than 600-fold, and pGDR18, which contains eighteen heptapeptide repeats, stimulated transcription more than 200-fold. Again, the level of transcription increased with an increase in the number of heptapeptide repeats of the CTD. The stimulation of transcription by the GAL4−CTD protein was

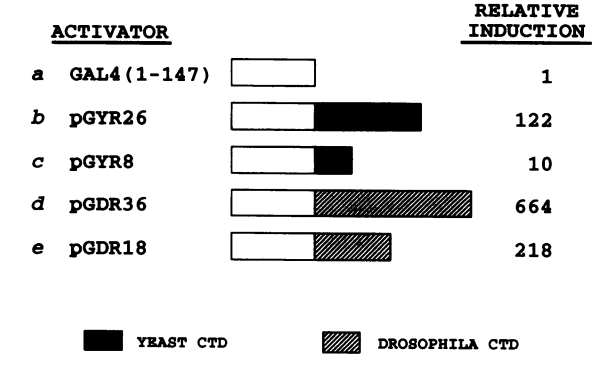

Figure 1. GAL4−CTD fusion proteins activate transcription in yeast. In plasmid pGYR26, the DNA-binding domain of GAL4 (amino acids 1−147) is fused to the entire CTD of yeast RNA polymerase II, and in pGYR8 to eight heptapeptide repeats from the CTD. Plasmid pGDR36 and pGDR18 contain GAL4 (1−147) fused, respectively, to thirty-six and eighteen heptapeptide repeats from the CTD of the *Drosophila* RNA polymerase II. These fusion constructs (see Materials and methods for details of construction) were introduced into the yeast strain, GGY1:171 which lacks the endogenous *GAL4* gene and harbors a *gal1−lacZ* reporter gene integrated at the *URA3* locus (19). β-galactosidase activity was determined by using the substrate CPRG as described (30). The values presented are averages of three to five independent assays, and are indicated as relative induction to that of GAL4 (1−147) after subtraction of background activity in the absence of GAL4. The relative levels of activity vary within 20%.

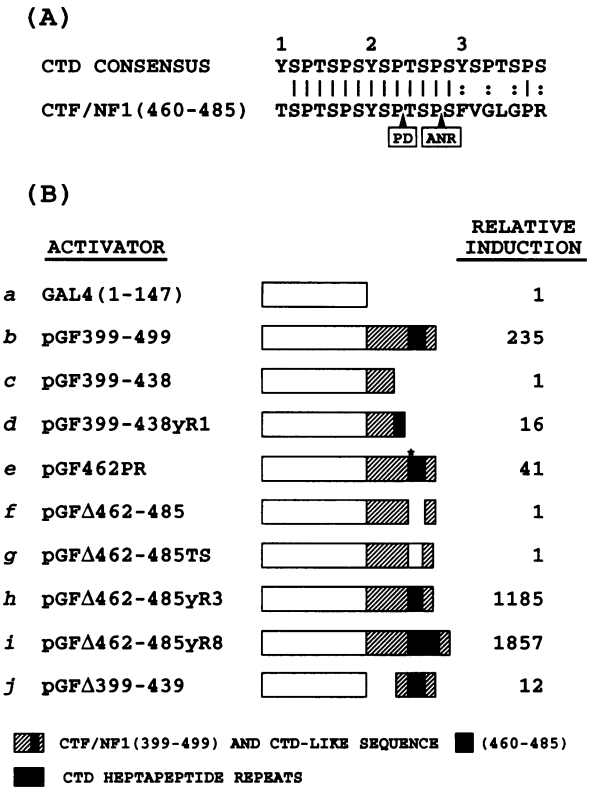

Figure 2. A CTD-like sequence in the proline-rich domain of CTF/NF1 is essential for the transcriptional activator function. (**A**) Sequence comparison between three repeats of the CTD consensus of RNA polymerase II and a sequence from position 460 to 485 of CTF/NF1. | designates amino acid identities, and : variations that are found at corresponding positions in multiple repeats of the native CTD sequences. Arrows indicate where two or three amino acids (boxed below the sequence) were removed in order to maximize the alignment. Such variations in the length of heptapeptide repeats are common in the RNA polymerase II CTD of *Drosophila* and other organisms (12, 13). (**B**) The CTD-like sequence is essential for the function of the proline-rich activation domain of CTF/NF1. The proline-rich domain (amino acids 399 to 499) of CTF/NF1 and its derivatives were fused to amino acids the GAL4 DNA-binding domain (amino acids 1–147, see Materials and methods for details of construction). The relative induction was determined as in Figure 1.

dependent on its binding to the upstream regulatory region: there was no stimulation of transcription from a *gal1–lacZ* reporter gene lacking the GAL4 binding sites (data not shown). These results demonstrate that both the yeast and *Drosophila* RNA polymerase II CTD can function upstream as a transcriptional activation domain when fused to the DNA-binding domain of a natural transcription factor.

The generality of the CTD acting upstream as an activation domain for transcription is further supported by a recent report that a GAL4–mouse CTD hybrid protein strongly stimulated transcription from a reporter gene in human Hela cells (36). Thus, the CTD of RNA polymerase II can function upstream as an activation domain for transcription in organisms as diverse as yeast and human. Since GAL4–CTD has the potential to interact directly with TFIID, these results are consistent with the view that activator proteins stimulate transcription by making direct contact with the general transcription machinery. More importantly, these results show that a domain of RNA polymerase II that is important for its recruitment to the promoter complex (16, 17, 37, 38) can also act upstream as an activation domain for transcription. This suggests a functional link between the recruitment of the RNA polymerase to a promoter and upstream activation of transcription.

The proline-rich domain of CTF/NF1 contains multiple CTD-like repeats that are essential for its transcriptional activating function

In order to assess the biological relevance of the CTD acting as a transcriptional activation domain, we searched for sequence similarities between the CTD of RNA polymerase II and known transcriptional activator proteins. The best match that we found is in the proline-rich transcriptional activation domain of the CCAAT-box-binding factor CTF/NF1 (26, 27). This domain (amino acids 399 to 499) contains a sequence with striking similarity to the heptapeptide repeat of the CTD. As shown in Figure 2A, this short sequence (amino acids 461 to 485) contains three CTD-like repeats. Amino acids 461 to 474 include one perfect heptapeptide repeat or twelve identities to two copies of the CTD consensus repeat if a bulge of two amino acids is allowed. As well, amino acids 478 to 485 partially match the next repeat when the natural variations in the CTD repeats are considered (12, 13). To test the importance of this CTD-like motif for the function of the proline-rich activation domain of CTF/NF1, we first tested the proline-rich domain (amino acids 399 to 499) for its ability to stimulate transcription in yeast. As shown in Figure 2B *b*, this domain, when fused to the DNA binding domain of GAL4, stimulated transcription from the *gal1–lacZ* reporter gene by over 200-fold. The stimulation of transcription by GAL4–CTF1 fusion proteins has also been observed from two other yeast promoters, *HIS3* (39) and *CYC1* (39, 40). Thus, the proline-rich domain of CTF/NF1 acts as a universal activator of transcription. In addition, expression of the GAL4–CTF1 fusion protein did not result in activation of a promoter without GAL4 binding sites (data not shown).

In order to assess the functional significance of the CTD-like motif in the proline-rich domain of CTF/NF1, we made a series of sequence alterations in the proline-rich activation domain. As shown in Figure 2B *c*, deletion of amino acids 439 to 499, which include the CTD-like motif, abolished the transcriptional activation function. Fusion of one and 5/7 heptapeptide repeats from the yeast CTD restored some activity (Figure 2B *d*; see further results below). After this manuscript was prepared for publication, an amino-terminal deletion analysis of the proline-rich domain was reported (41). Although both of these results appear to suggest a role of the CTD-like motif for the transcriptional activity of the proline-rich domain, the interpretation of these results is limited by the low resolution of such one-directional deletion analyses.

To test directly the importance of the CTD-like motif for the transcriptional activator function of CTF/NF1, we performed site-directed mutagenesis on the CTD-like motif. A point mutation of the conserved proline residue at position 462 (pGF462PR) reduced the level of transcriptional activity by six-fold (Figure 2B *e*), and an internal deletion of amino acids 462 to 485 that removed only the CTD-like motif (pGFΔ462-485) completely eliminated the transcriptional activity of the proline-rich activation domain (Figure 2B *f*). This demonstrates that the CTD-like motif is essential for the transcriptional activation function of the proline-rich domain of CTF/NF1. We then inserted into this deletion either a sequence encoding 21 unrelated amino acids (Figure 2B *g*) or a sequence encoding three CTD heptapeptide repeats (Figure 2B *h*). The insertion of 21 unrelated amino acids

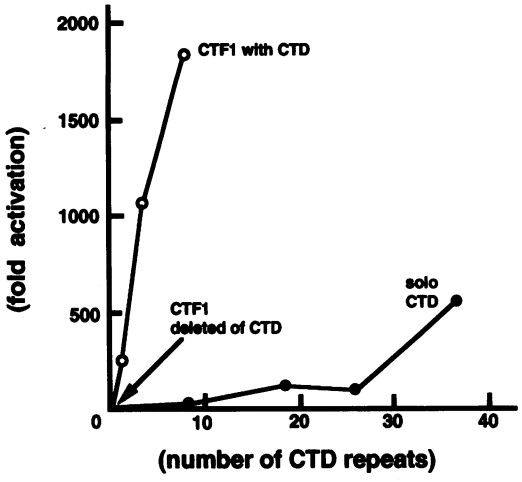

Figure 3. A graphical presentation of results from Figures 1 and 2 showing that at least two separable components in the proline-rich domain of CTF/NF1 are required for its optimal function. The proline-rich domain of CTF/NF1, when deleted of the CTD-like motif, showed no transcriptional activity, while the CTD heptapeptide repeats stimulated transcription to a much higher level when present in the context of the proline-rich domain of CTF/NF1 than when acting alone. Thus, the CTD heptapeptide repeats and the other component(s) of the proline-rich domain of CTF/NF1 act synergistically to stimulate transcription.

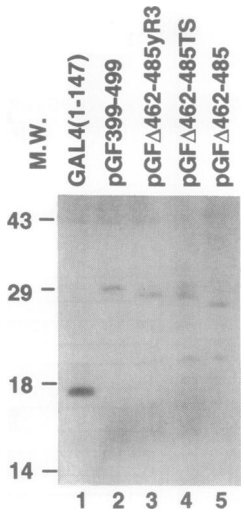

Figure 4. Protein blot analysis using antibody against the GAL4 DNA-binding domain (amino acids 1–147). Protein extracts were prepared from yeast cells transformed with plasmids containing wild type or mutant derivatives of the proline-rich activation domain of CTF/NF1, which are fused to the DNA-binding domain of GAL4 (see Figure 2). All the samples contained equal amounts of proteins (as predetermined by Coomassie blue staining). The molecular markers (M.W.) are in kd. Lane 1, GAL4 (1–147); lane 2, pGF399-499; lane 3, pGFΔ462-485yR3; lane 4, pGFΔ462-485TS; and lane 5, pGFΔ462-485.

(pGFΔ462-485TS) conferred no detectable transcriptional activity above background levels, while the insertion of three CTD heptapeptide repeats (pGFΔ462-485yR3) rescued the functional defect of the mutant derivative of the proline-rich domain. Note also that substitution of three perfect heptapeptide repeats for the less conserved, naturally occurring, CTD-like motif conferred a five-fold higher level of transcriptional activation than the wild type proline-rich domain (compare Figure 2B *h* and *b*), and substitution of eight heptapeptide repeats from the yeast RNA polymerase II for the CTD-like motif conferred an eight-fold higher level of transcriptional activation than the wild type proline-rich domain (compare Figure 2B *h* and *j*). Thus, the level of activation is roughly proportional to the number of CTD repeats. These results further demonstrate that the conserved heptapeptide repeats are a critical component for the transcriptional activation function of the proline-rich domain of CTF/NF1.

It should be noted, however, that the CTD-like motif may not be the only functional component of the proline-rich activation domain of CTF/NF1, since the same number of heptapeptide repeats conferred a much higher level of transcription when inserted into the proline-rich domain of CTF/NF1 than when acting alone (compare Figure 1 *c* with Figure 2B *i*). The other component(s) of the proline-rich domain of CTF/NF1 apparently has no independent activity (Figure 2B *c* and *f*), but contribute strongly to the activity of the heptapeptide repeats. We noticed that in addition to the essential CTD-like motif, amino acids 421 to 446 of the proline-rich domain of CTF/NF1 include bulky hydrophobic residues which are distributed in a pattern similar to those in the VP16 and other activation domains (42, 43). To assess the importance of these hydrophobic residues for the transcriptional activity of the proline-rich domain of CTF/NF1, we made a deletion of amino acids 399 to 439 to remove most of these hydrophobic residues. As shown in Figure 2B *j*, this deletion dramatically reduced, though did not eliminate, the transcriptional activity of the proline-rich domain of CTF/NF1. Thus, the CTD heptapeptide repeats and this other component(s) in the proline-rich domain of CTF/NF1 act synergistically to stimulate transcription (see Figure 3 for a graphical description of the functional relationship between the two components).

Protein blot analysis using antibody against the GAL4 DNA-binding domain showed that fusion proteins containing both the wild-type proline-rich domain and the mutant derivatives were produced in similar amounts in yeast cells (Figure 4). Thus, the inability of the mutants to stimulate transcription is not a result of instability or underproduction of the mutant proteins. This further supports the conclusion that the CTD-like motif is essential for the activator function of the proline-rich domain of CTF/NF1.

The proline-rich activation domain of CTF/NF1 interacts directly with the TATA-box-binding protein

It has been shown previously that the heptapeptide repeat of the CTD interacts directly with TBP, the TATA-box-binding subunit of the general transcription factor TFIID (16, 17). This suggests that the proline-rich activation domain of CTF/NF1 might also interact directly with TBP. To test this possibility, we constructed a fusion protein that contains the proline-rich activation domain of CTF/NF1 fused to the *E.coli* maltose-binding protein (MBP). We then expressed the MBP/CTF1 fusion protein in *E.coli*, and purified the fusion protein by affinity chromatography. We passed this fusion protein through a column prebound with a GST–yeast TBP fusion protein or a control column prebound with only the GST protein. As shown in Figure 5A, the intact MBP/CTF1 fusion protein (indicated by an arrow in lane 1) was preferentially retained, compared to the smaller degradation products and MBP, on the GST–yeast TBP column (lane 2), but not on the control column (lane 3). This result demonstrates that the proline-rich activation domain of CTF/NF1 can interact directly with TBP.

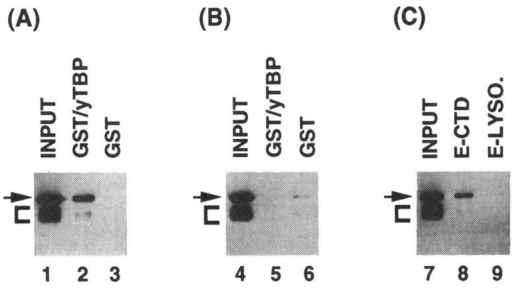

Figure 5. The CTD-like motif of the proline-rich domain of CTF/NF1 is important for its interaction with TBP. The wild type proline-rich domain (amino acids 399–499) of CTF/NF1 and a mutant derivative (CTFΔ462-485TS) were fused to the *E.coli* maltose-binding protein (MBP). The fusion proteins were expressed in and partially purified from *E.coli*. (**A**) The partially purifed MBP–CTF1 fusion protein (lane 1) was loaded onto a column prebound with a GST–yeast TBP fusion protein (lane 2) or a control column prebound with only the GST (lane 3). Elutions from these columns were analyzed by protein blotting using an anti-MBP antibody. Arrow indicates the intact MBP–CTF1 fusion protein, and bracket includes degradation products and MBP. (**B**) An MBP fusion protein with the mutant derivative was analyzed as in (A). (**C**) A synthetic CTD peptide competes with CTF1 for interaction with TBP. Partially purified MBP–CTF1 fusion protein (lane 7) was loaded onto two identical columns prebound with the GST–yeast TBP fusion protein. Following a wash with affinity chromatography buffer (ACB) containing 50 mM KCl, the bound proteins were eluted either with ACB/50 mM KCl/1.5 mg ml^{-} of the synthetic CTD peptide (lane 8) or with ACB/50 mM KCl/1.5 mg ml^{-} of the egg white lysozyme as a control (lane 9), and analyzed by protein blotting as in (A).

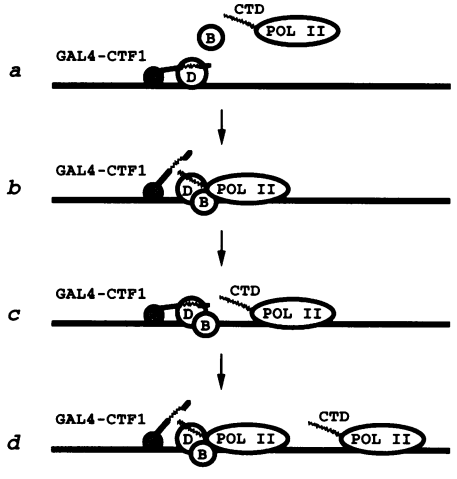

Figure 6. A tether and competition model of transcriptional activation. For simplicity, only GAL4–CTF1, TFIID (D), TFIIB (B), and RNA polymerase II (POL II) are shown. The activator GAL4–CTF1 could act at both pre-initiation and post-initiation steps to stimulate transcription. Direct interaction of the CTD-like sequence in the proline rich domain of CTF1 with TFIID may first recruit the general factor TFIID to the promoter (*a*) or block its interactions with negative regulatory factors, thus facilitating the subsequent assembly of a transcription complex (*b*). Following the assembly of the transcription complex, the potential competitive interaction of the upstream-bound CTD-like sequence with TFIID may displace or destabilize the interaction of the RNA polymerase II CTD with TFIID, thus accelerating the release of the RNA polymerase from the basal promoter complex (*c*). For multiple rounds of transcription, the relative rate constants for association with and dissociation from TFIID of the upstream-bound CTD-like sequence and of the CTD of RNA polymerase II may be a factor in determining the frequency at which RNA polymerase II is recruited to and subsequently released from the basal promoter complex (*b* to *d*).

Since the derivative (CTFΔ462-485TS) containing a substitution of the CTD-like motif with unrelated amino acids showed very little transcriptional activity (Figure 2B *g*), we decided to determine the effect of the substitution on the binding of the mutant protein to TBP. Again, we constructed, expressed in, and purified from *E.coli* a fusion protein that contains the mutant derivative fused to MBP. We then passed this fusion protein through a column prebound with a GST–yeast TBP fusion protein or a control column prebound with only the GST protein. As shown in Figure 5B, there was no specific retention of the mutant protein on the GST/yTBP column (lanes 5 and 6). This result demonstrates that the CTD-like motif is important for the interaction of the proline-rich domain with TBP, and further suggests that the affinity of the proline-rich domain for TBP is important for its *in vivo* transcriptional activity.

To demonstrate further the specificity of interaction of the CTD-like motif with TBP and its relationship to natural CTD repeats, we synthesized a CTD heptapeptide, and tested the synthetic peptide for its ability to compete with the proline-rich domain of CTF/NF1 for binding to TBP. To do this, the MBP–CTF1 fusion protein prebound to a GST–yeast TBP column was eluted either with the synthetic CTD heptapeptide or with the egg-white lysozyme (Boehringer Mannheim) as a control. As shown in Figure 5C, the intact MBP/CTF1 fusion protein was preferentially eluted by the CTD heptapeptide (lane 8), but not by lysozyme (lane 9). This result indicates that both the CTD-like motif of CTF/NF1 and the heptapeptide repeats of the RNA polymerase II CTD interact with a similar or identical target on the TBP protein.

DISCUSSION

We have demonstrated that the heptapeptide repeats that constitute the carboxy-terminal domain of RNA polymerase II can function to stimulate transcription when fused to the DNA binding domain of GAL4. More importantly, we have shown that the CTD-like heptapeptide repeats are essential functional components of a natural activator protein, the CCAAT-box binding factor CTF/NF1. The heptapeptide repeats may act through a number of potential mechanisms to stimulate transcription from an upstream site. The CTD of RNA polymerase II has been shown to interact directly with the TATA-binding protein, TBP (16, 17). We have shown in this study that the CTD-like repeats in the proline-rich activation domain of CTF/NF1 are also important for its interaction with TBP. Thus, one can envision a simple model in which this interaction may recruit the general transcription factor TFIID and stabilize its binding to the promoter (Figure 6 *a*), and/or block its interaction with negative regulatory factors. This would facilitate the recruitment of TFIIB and subsequent assembly of a transcription complex (Figure 6 *b*), thereby leading to enhanced transcription in ways that have previously been proposed for acidic and other activators (8–11, 44–52). Since in the current study we did not test the ability of CTF/NF1 to interact with other general factors, it remains a formal possibility that CTF/NF1 may interact with other general factor(s) such as TFIIB, thus directly recruiting it to the promoter.

The CTD of RNA polymerase II has been implicated in the recruitment of the RNA polymerase into the transcription complex through its direct interaction with the general transcription factor TFIID (16, 17, 37, 38). However, association of the CTD with TFIID could prevent the RNA polymerase from leaving the complex, as has been proposed (17, 53–55). This might also

be the case for genes such as *Drosophila hsp70* and human *c-myc*, where the escape of a transcriptionally-engaged RNA polymerase II molecule from an early elongation pause appears to be rate-limiting (56–61). If in such cases the CTD–TFIID contact were indeed responsible for holding back the polymerase near the promoter, disruption of this contact would result in transcriptional activation. Several groups have suggested that phosphorylation of the CTD may lead to disruption of the CTD–TFIID contact, and thus the release of the RNA polymerase from the initiation complex (53–55, 62, 63). This would allow the RNA polymerase to proceed into productive elongation. As an alternative to this model, we suggest that, as a result of competition for interaction with TFIID (see Figure 5C), the upstream-bound activator CTD or the proline-rich activation domain of CTF/NF1 could have the potential to displace or destabilize the contact of the CTD of RNA polymerase II with TFIID, accelerating the release of RNA polymerase II from an initiation or early elongation complex (Figure 6 *c*). Thus, our results would imply that an activator such as GAL4–CTD could have the potential to act at both pre-initiation and post-initiation steps to stimulate transcription (Figure 6).

One may argue that a CTD or CTD-like sequence bound near the promoter may reduce the rate at which RNA polymerase II enters the promoter complex, since it could compete with the CTD of the polymerase for interaction with TFIID (see for example, Figure 5C). We suggest that the rate of transcription following the initial recruitment of TFIID to a promoter would be dependent, at least in part, on the relative rate constants for association with and dissociation from TFIID of the upstream-bound activator CTD and the RNA polymerase CTD. In this context, it is conceivable that specific rate constants, and thus the level of transcription, can be modulated by other factors such as SRB2 and SIN1 that may interact with the CTD (16, 38, 64).

We have focused our discussion on two possible steps at which the upstream-bound activator CTD may act. However, the activator CTD might have an effect(s) on other steps as well. For example, when bound to DNA the CTD might have an effect on local chromatin (65), or it might interact with histones or other chromosomal proteins as has been proposed (34). This might in turn affect the rate of transcription (66–68).

Our study of the CTD of RNA polymerase II has led to the finding that this conserved domain of the RNA polymerase can act upstream as an activation domain of transcription, and that the transcriptional activation domain of a natural transcription factor possesses copies of the heptapeptide repeat of the CTD that are essential for its activation function. This is intriguing because a common structural component is used on the one hand by the RNA polymerase in the recognition of a promoter complex through its interaction with TBP (16, 17, 38), and on the other by the proline-rich activator of CTF/NF1 in the activation of transcription. We showed that the CTD-like motif is important both for the transcriptional activation function of the proline-rich domain of CTF/NF1 and for its interaction with TBP. We further showed that a synthetic CTD peptide can compete with the proline-rich activation domain of CTF/NF1 for interaction with TBP. These results strongly suggest that RNA polymerase II and the activation domain of CTF/NF1 can both interact with the same target in the general transcription machinery, possibly in a competitive manner during and after the assembly of a transcription complex (see discussion above). This would provide a simple mechanistic model for the future study of transcriptional activation by upstream activator proteins in eukaryotes.

ACKNOWLEDGEMENTS

We thank Drs J.Archambault, R.Tjian, M.Ptashne, A.Weil, and A.Greenleaf for providing strains and plasmids; Dr M.Ptashne for the antibody against the GAL4 DNA-binding domain; Drs E.Keller and M.Wolfner for critically reading this manuscript, and members of the Friesen and Lis laboratories for discussion and comments; J.Li for help in protein affinity chromatography experiments; J.Ingraffea for assistance in purification of GST fusion proteins and Dr J.Peng at Cornell Biotechnology Facility for the synthesis of the CTD peptide. This work was supported by grants from MRC and NCI (Canada) to J.D.F and J.G., and from NIH (USA) to J.T.L. H.X. was a recipient of a postdoctoral fellowship from the Medical Research Council of Canada.

REFERENCES

1. Buratowski, S., Hahn, S., Guarente, L. and Sharp, P. A. (1989) *Cell* **56:** 549–561.
2. Burton, Z. F., Ortolan, L. G. and Greenblatt, J. (1986) *EMBO J.* **5:** 2923–2930.
3. Sawadogo, M. and Sentenac, A. (1990) *Annu. Rev. Biochem.* **59:** 711–754.
4. Zawel, L. and Reinberg, D. (1993) *Prog. Nucl. Acid. Res. Mol. Biol.* **44:** 67–108.
5. Mitchell, P. J. and Tjian, R. (1989) *Science* **245:** 371–378.
6. Ptashne, M. (1988) *Nature* **335:** 683–689.
7. Ptashne, M. and Gann, R. A. (1990) *Nature* **346:** 329–331.
8. Ingles, C. J., Shales, M., Cress, W. D., Triezenberg, S. J. and Greenblatt, J. (1991) *Nature* **351:** 588–590.
9. Lin, Y. S. and Green, M. R. (1991) *Cell* **64:** 871–981.
10. Lin, Y. S., Ha, I., Maldonado, E., Reinberg, D. and Green, M. R. (1991) *Nature* **353:** 569–571.
11. Stringer, K. F., Ingles, C. J. and Greenblatt, J. (1990) *Nature* **345:** 783–786.
12. Corden, J. L. (1990) *Trends Biochem. Sci.* **15:** 383–387.
13. Young, R. A. (1991) *Annu. Rev. Biochem.* **60:** 689–715.
14. Scafe, C., Chao, D., Lopes, J., Hirsch, J. P., Henry, S. and Young, R. A. (1990) *Nature* **347:** 491–494.
15. Allison, L. A. and Ingles, C. J. (1989) *Proc. Natl. Acad. Sci. USA* **86:** 2794–2798.
16. Thompson, C. M., Koleske, A. J., Chao, D. M. and Young, R. A. (1993) *Cell* **73:** 1361–1375.
17. Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D. and Aloni, Y. (1992) *Cell* **69:** 871–881.
18. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (eds.) (1987–1989) Current Protocols in Molecular Biology-Laboratory Manuals. John Wiley & Sons, Inc., Vol. 2.
19. Gill, G. and Ptashne, M. (1987) *Cell* **51:** 121–126.
20. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
21. Archambault, J., Drebot, M. A., Stone, J. C. and Friesen, J. D. (1992) *Mol. Gen. Genet.* **232:** 408–414.
22. Ma, J. and Ptashne, M. (1987) *Cell* **51:** 113–119.
23. Sorger, P. K. and Pelham, H. R. B. (1988) *Cell* **54:** 855–864.
24. Sikorski, R. S. and Hieter, P. (1989) *Genetics* **122:** 19–27.
25. Jokerst, R. S., Weeks, J. R., Zehring, W. A. and Greenleaf, A. L. (1989) *Mol. Gen. Genet.* **215:** 266–275.
26. Mermod, N., O'Neill, E. A., Kelly, T. J. and Tjian, R. (1989) *Cell* **58:** 741–753.
27. Santoro, C., Mermod, N., Andrews, P. C. and Tjian, R. (1988) *Nature* **334:** 218–224.
28. Hill, J., Ian, K. A., Donald, G. and Griffiths, D. E. (1991) *Nucleic Acids Res.* **19:** 5791.
29. Guarente, L. and Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* **78:** 2199–2203.
30. Simon, J. A. and Lis, J. T. (1987) *Nucleic Acids Res.* **15:** 2971–2988.
31. Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M. and Ingles, C. J. (1988) Mol. Cell. Biol. **8:** 321–329.
32. Bartolomei, M. S., Halden, N. F., Cullen, C. R. and Corden, J. L. (1988) *Mol. Cell. Biol.* **8:** 330–339.
33. Liao, S.-M., Taylor, I. C. A., Kingston, R. E. and Young, R. A. (1991) *Genes Dev.* **5:** 2431–2440.

34. Nonet, M., Sweetser, D. and Young, R. A. (1987) *Cell* **50:** 909–915.
35. Zehring, W. A., Lee, J. M., Weeks, J. R., Jokerst, R. S. and Greenleaf, A. L. (1988) *Proc. Natl. Acad. Sci. USA* **85:** 3698–3702.
36. Seipel, K., Georgiev, O., Gerber, H.-P. and Schaffner, W. (1993) *Nucle. Acids Res.* **21:** 5609–5615.
37. Conaway, R. C., Bradsher, J. N. and Conaway, J. W. (1992) *J. Biol. Chem.* **267:** 8464–8467.
38. Koleske, A. J., Buratowski, S., Nonet, M. and Young, R. A. (1992) *Cell* **69:** 883–894.
39. Xiao, H. (unpublished results)
40. Kim, T. K. and Roeder, R. G. (1993) *J. Biol. Chem.* **268:** 20866–20869.
41. Kim, T. K. and Roeder, R. G. (1993) *Nucle. Acid Res.* **22:** 251.
42. Cress, W. D. and Triezenberg, S. J. (1991) *Science* **251:** 87–90.
43. Regier, J. L., Shen, F. and Triezenberg, S. J. (1993) *Proc. Natl. Acad. Sci. USA* **90:** 883–887.
44. Choy, B. and Green, M. R. (1993) *Nature* **366:** 531–536
45. Greenblatt, J. (1991) *Cell* **66:** 1067–1070.
46. Hawley, D. (1991) *TIBS* **16:** 317–318.
47. Horikoshi, N., Maguire, K., Kralli, A., Maldonado, E., Reinberg, D. and Weinmann, R. (1992) *Proc. Natl. Acad. Sci. USA* **88:** 5124–5128.
48. Lee, W. S., Kao, C. C., Bryant, G. O., Liu, X. and Berk, A. J. (1991) *Cell* **67:** 365–376.
49. Meisterernst, M. and Roeder, R. G. (1991) *Cell* **67:** 557–567.
50. Sharp, P. A. (1991) *Nature* **351:** 16–18.
51. Sundseth, R. and Hansen, U. (1992) *J. Biol. Chem.* **267:** 7845–7855.
52. Wang, W., Gralla, J. D. and Carey, M. (1992) *Genes Dev.* **6:** 1716–1727.
53. Chesnut, J. D., Stephens, J. H. and Dahmus, M. E. (1992) *J. Biol. Chem.* **267:** 10500–10506.
54. Dahmus, M. E. and Dynan, W. S. (1992) In McKnight, S. and Yamamoto, K. (eds.), Transcriptional Regulation. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
55. Lu, H., Flores, O., Weinmann, R. and Reinberg, D. (1991) *Proc. Natl. Acad. Sci. USA* **88:** 10004–10008.
56. Giardina, C., Perez-Riba, M. and Lis, J. T. (1992) *Genes Dev.* **6:** 2190–2200.
57. Krumm, A., Meulia, T., Brunvand, M. and Groudine, M. (1992) *Genes Dev.* **6:** 2201–2213.
58. O'Brien, T. and Lis, J. T. (1991) *Mol. Cell. Biol.* **11:** 5285–5290.
59. Rougvie, A. E. and Lis, J. T. (1990) *Mol. Cell. Biol.* **10:** 6041–6045.
60. Rougvie, A. E. and Lis, J. T. (1988) *Cell* **54:** 795–804.
61. Strobl, L. J. and Eick, D. (1992) *EMBO J.* **11:** 3307–3314.
62. Lis, J. and Wu, C. (1993) *Cell* **74:** 1–4.
63. Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M. and Greenleaf, A. L. (1993) *Genes Dev.* **7:** 2329–2344.
64. Peterson, C. L., Kruger, W. and Herskowitz, I. (1991) *Cell* **64:** 1135–1143.
65. Suzuki, M. (1990) *Nature* **344:** 362–365.
66. Dusserre, Y. and Mermod, N. (1992) *Mol. Cell. Biol.* **12:** 5228–5237.
67. Laybourn, P. J. and Kadonaga, J. T. (1991) *Science* **254:** 238–245.
68. Workman, J. L., Taylor, I. C. A. and Kingston, R. E. (1991) *Cell* **64:** 533–544.