MOLECULAR AND CELLULAR BIOLOGY, Nov. 1992, p. 5094–5101
0270-7306/92/115094-08$02.00/0

Vol. 12, No. 11

Definition of the Transcriptional Activation Domain of Recombinant 43-Kilodalton USF

BERNHARD J. KIRSCHBAUM, PHILIPPE POGNONEC, AND ROBERT G. ROEDER*

The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399

Received 10 July 1992/Returned for modification 6 August 1992/Accepted 13 August 1992

The cellular transcription factor USF is involved in the regulation of both cellular and viral genes and consists of 43- and 44-kDa polypeptides which independently show site-specific DNA binding. Cloning of the corresponding cDNA revealed that the 43-kDa polypeptide (USF^{43}) is a member of the basic (B)-helix-loop-helix (HLH)-leucine zipper (LZ) family of proteins and provided a means for its functional dissection. Initial structure-function studies revealed that the HLH and LZ regions are both important for USF^{43} oligomerization and DNA binding. The studies presented here have focused on the determination of domains that contribute to transcriptional activation in vitro and show that (i) both a small region close to the N terminus and a region between residues 93 and 156 contribute strongly to transcriptional activation, (ii) full activation depends on the presence of both domains, (iii) the B-HLH-LZ region has no intrinsic activation potential but DNA binding is absolutely required for transcriptional activation, and (iv) the B-HLH-LZ region can be replaced by the Gal4 DNA binding domain without loss of activation potential.

The ubiquitous cellular transcription factor USF is a member of the helix-loop-helix (HLH) family of regulatory proteins. USF was originally described as an upstream stimulatory factor that binds to the core sequence CACGTG in the adenovirus major late promoter (5, 9, 22, 23, 31). It subsequently was shown to be involved in the regulation of a number of cellular genes, including those encoding mouse metallothionein I (6), rat γ-fibrinogen (8), human growth hormone (25), and *Xenopus* TFIIIA (15, 33). USF was also shown to interact with the essential μE3 motif in the immunoglobulin heavy-chain enhancer (1) as well as with a pyrimidine-rich initiator element found in many genes (30). Together with the further demonstration of cooperative interactions between USF and the initiator-binding protein TFII-I (30), these findings raised the possibility of a more general involvement of USF in transcriptional regulation.

Purified human USF is composed of 43- and 44-kDa polypeptides (USF^{43} and USF^{44}) which show independent site-specific DNA binding (32). Cloning and characterization of cDNAs encoding the 43-kDa component was first described for human (12) and subsequently for *Xenopus* (15) and sea urchin (16) proteins. Together with the Myc proteins (for a review, see reference 19), AP4 (13), TFE3 (2), and TFEB (4), human USF^{43} falls into a family containing basic region (B), HLH, and leucine zipper (LZ) motifs. A comparison of human and *Xenopus* cDNAs revealed high sequence conservation almost throughout the entire protein (15). These USF isoforms possess about 80% sequence identity at the amino acid level, with only a 40-amino-acid region in the midportion of the protein showing strong divergence. In contrast, sea urchin USF shows strong divergence from human and *Xenopus* USF in large portions N terminal of the B-HLH domain and, surprisingly, lacks the C-terminal LZ motif. Mutagenesis studies on human USF^{43} pointed to the importance of both the HLH and LR regions for oligomerization and DNA binding (12). A distinct type of regulation is achieved via two cysteine residues within the HLH region (26), the formation of intra- and intermolecular disulfide bridges resulting in an inhibition of DNA binding and transcriptional activation. Recombinant USF^{43} ($rUSF^{43}$) binds to DNA and activates transcription in crude reconstituted systems in a manner indistinguishable from that of natural USF^{43} and USF^{44} (28) and thus can be used for further functional dissection.

The objective of this study was to determine the activation domain(s) within human USF^{43}. A major obstacle to an analysis of this question in vivo is the high level of endogenous USF in most or all cell types, making the results of transfection experiments difficult to interpret. To circumvent this problem, we used an in vitro transcription system which consisted of nuclear extracts that were immunodepleted of endogenous USF and showed only basal-level transcription (28). In this system, exogenous $rUSF^{43}$ elevated transcription from a USF-responsive promoter (a human immunodeficiency virus [HIV] core promoter with two upstream USF sites) to the level observed in unfractionated nuclear extracts. The analysis of mutated forms of USF^{43} allowed the identification of two regions N terminal of the B-HLH-LZ region that contribute to transcriptional activation.

MATERIALS AND METHODS

Construction of recombinant plasmids. (i) USF expression vectors. USF recombinants were expressed in the bacterial expression system developed by Studier et al. (34) (Novagen). Deletion mutants containing only parts of the first 196 amino acids of $rUSF^{43}$ (total length, 311 amino acids) were prepared by ligation of polymerase chain reaction (PCR)-derived fragments or specific adaptors, respectively, into convenient restriction sites of the original full-length pET3d/USF construct (28). The *Nco*I site in the vector (encoding the first ATG) provided the 5′ cloning site. The single *Kpn*I, *Avr*II, and *Ava*I sites in the USF coding sequence were used as 3′ cloning sites. PCR reactions using the full-length pET3d/USF construct as a template were carried out with Pfu (Stratagene) as instructed by the supplier. The following oligonucleotides were designed as upstream PCR primers for N-terminal deletions (the *Nco*I restriction sites are underlined): Δ1-15, 5′-TCGCCATGGTGCAGATTCAGGAAGG

* Corresponding author.

T-3'; Δ1-39, 5'-TCGCCATGGCTGCCACCTTCCCTGAC-3'; Δ1-59, 5'-TCGCCATGGTGATGTACAGGGTGATC-3'; Δ1-80, 5'-TCGCCATGGCCATCAGTGGCTACCCT-3'; and Δ1-93, 5'-TCGCCATGGCGGTGATCCAGGGTGCT-3'. An oligonucleotide containing a *Kpn*I site (5'-TCGGGTAC CCCCTGCCCCATCTCC-3') was used as a downstream partner. A larger N-terminal deletion (Δ1-156) was created by the combination of 5'-TCGCCATGGGCACTGGTC AATTCTTTGTG-3' containing an *Nco*I site and a primer hybridizing far downstream in the USF coding sequence (covering amino acids [aa] 272 to 266). This PCR product included the single *Avr*II site in the USF cDNA as a 3' cloning site.

To create internal deletions, the *Kpn*I site in the USF cDNA (bp 390) was sequentially moved toward the start codon. This resulted in internal deletions of increasing sizes starting from aa 130 and proceeding toward the N terminus of the protein. The following oligonucleotides were used with the T7 primer (hybridizing in the vector) as upstream partners (the *Kpn*I restriction site is underlined): Δ105-130, 5'-TCGGGTACCATCATCACTGGTGAAAGC-3'; Δ94-130, 5'-TCGGGTACCCTGGGTCATGGATTGAGT-3'; Δ80-130; 5'-TCGGGTACCAGTTCCCTCAGTTTGG-3'; Δ58-130, 5'-TCGGGTACCCCCATTCTCAGTTCGG-3'; and Δ26-130, 5'-TCGGGTACCAGTAGCCACTGCACCT-3'.

Aliquots of these PCR reaction mixtures were extracted once with chloroform, ethanol precipitated, and subsequently cut with the appropriate restriction enzymes. The resulting products were purified on 1.5% low-melting-point agarose gels and ligated into the correspondingly prepared plasmid pET3d/USF.

Some of the mutants were produced by connecting two restriction sites with a specifically designed adaptor. These phosphorylated adaptors contained an *Nco*I site at their 5' ends and a *Kpn*I (Δ1-130), *Avr*II (Δ1-175), or *Ava*I (Δ1-196) site, respectively, at their 3' ends. The adaptor sequences for mutants Δ1-130 and Δ1-175 are as follows:

```
Δ1-130, 5'-CATGACACCACCAGTAC-3'
              3'-TGTGGTGGT-5'
Δ1-175, 5'-CATGCAGCGCTCAATTGCCC-3'
              3'-GTCGCGAGTTAACGGGGATC-5'
```

For the construction of Δ1-196, the USF cDNA was cloned into the *Eco*RI site of pBSK$^-$ to obtain a single-strand copy of the noncoding strand. By using the procedure of Kunkel et al. (17), an *Nco*I site was then introduced by directed mutagenesis with the following oligonucleotide (the *Nco*I site is underlined): 5'-CCCCGGACGACCATGGATGAG AAACGC-3'. The *Nco*I-*Bam*HI fragment was then subcloned into the corresponding sites of pET3d.

After ligation and transformation into *Escherichia coli* DH5α, individual colonies were screened by PCR with proper primer combinations, using 3-μl aliquots of bacterial minicultures as a template. The recombinant DNA was isolated by minipreparations, applying the alkaline lysis method. Aliquots of these DNA preparations were used for sequencing. The remaining DNA was introduced into BL21 cells (Novagen), allowing expression of the recombinant proteins.

(ii) Gal4 expression vectors. Plasmid pD10Gal4(1-94) and bacterial strain MC1061/PDMI,1 for expression of H6Gal4(1-94) were generously provided by C. Rosen. This expression system is derived from the pDS expression plasmid (7, 10, 35). Three different USF domains (aa 17 to 104, 131 to 196, and 17 to 196) were cloned in frame into the single *Bam*HI and *Xba*I sites 3' of the pD10Gal4(1-94) coding sequence. The USF fragments with the appropriate restriction sites at their ends were produced by PCR with Pfu (Stratagene) as described above, using the following oligonucleotides as primers (the 5' *Bam*HI and the 3' *Xba*I cloning sites are underlined): a, 5'-TCGGGATCCAGATTCAGGAAGGTGC A-3'; b, 5'-TCGTCTAGAAATCATCACTGGTGAAAGC-3'; c, 5'-TCGGGATCCAGGGTACCACATCGGGGAGTAC A-3'; and d, 5'-TCGTCTAGAAAGTCGTCCGGGGAGCTT CTGA-3'. The combination of primers a and b created USF domain aa 17 to 104, the combination of primers c and d produced USF domain aa 131 to 196, and the combination of primers a and d created USF domain aa 17 to 196. The constructs were raised in MC1061/PDMI,1 cells (supplied by C. Rosen). Plasmid pDMI,1 allows overproduction of the *lac* repressor (important for stable maintenance of pDS-derived vectors), confers kanamycin resistance, and contains the p15A replicon (7).

Expression and purification of recombinant proteins. For induction of the pET3d constructs, 50 ml of M9ZB (M9 minimal medium supplemented with 10 g of N-Z-amine A and 5 g of NaCl per liter)-ampicillin (final concentration of 500 mg/ml) was inoculated with 2 ml of an overnight culture grown at 30°C. Induction of the recombinant protein was started at an optical density of 580 nm of 0.6 (usually 3 to 4 h after inoculation) by addition of isopropylthiogalactopyranoside (IPTG; final concentration of 0.5 mM). For induction of the Gal4 constructs, 50 ml of LB supplemented with ampicillin and kanamycin (final concentrations of 100 and 50 μg/ml, respectively) were inoculated with 2 ml of an overnight culture. Induction was started at an optical density at 580 nm of 0.4 by addition of 1 mM (final concentration) IPTG. At the same time $Zn(acetate)_2$ was added at a final concentration of 100 μM. Induction was continued for 3 to 4 h at 30°C. For direct electrophoretic analysis, aliquots of the culture were removed before (1.5 ml) and after (0.5 ml) induction. These samples were briefly centrifuged, and the bacterial pellets were directly dissolved in Laemmli loading buffer. Sonication was usually necessary to reduce the viscosity prior to electrophoresis. The remaining bacteria were pelleted and stored at −20°C until further processed.

The initial purification of proteins without histidine-tag (artificial leader sequence of six histidine residues) was done by the ammonium sulfate method described before (27). The recombinant proteins were further purified by passage through a conventional heparin-Sepharose column and, as a final step, DEAE 5 PW high-pressure liquid chromatography (HPLC). Protein preparations were routinely checked by sodium dodecyl sulfate (SDS)-gel electrophoresis and Coomassie staining.

Histidine-tag-containing Gal4/USF fusion proteins were batch purified on an Ni^{2+} resin. The pellet of a 50-ml culture was resuspended in 1 ml of lysis buffer (20 mM Tris-HCl [pH 7.9], 10% glycerol, 500 mM KCl, 10 μM $ZnCl_2$, 1 mM phenylmethylsulfonyl fluoride, 1% [vol/vol] aprotinin, 5 μg of leupeptin per ml, 10 mM dithiothreitol, 0.1% [vol/vol] Nonidet P-40), incubated for 10 min on ice-salt water, and sonicated until the viscosity dropped (usually 15 s total). After centrifugation, the cell debris was discarded and the supernatant was mixed with 100 μl of Ni^{2+} resin (equilibrated in lysis buffer). The tubes were put on ice-salt water for 2 h. During this time the Ni^{2+} resin was kept suspended by regularly inverting the tubes. After centrifugation, the supernatant was carefully removed and the Ni^{2+} resin was washed five times with lysis buffer (1 ml) and once with lysis buffer containing 20 mM imidazole. The bound protein was eluted by addition of 0.8 ml of lysis buffer containing 100 mM

imidazole. The eluate was diluted 1:5 with lysis buffer to reduce the imidazole concentration to 20 mM, and the purification was repeated once more. The protein was eluted with 400 μl of BC100 (20% glycerol, 100 mM KCl, 20 mM Tris-Cl [pH 7.4]) containing 100 mM imidazole and 10 μM $ZnCl_2$, dialyzed against BC100 containing 10 μM $ZnCl_2$, and stored at −70°C.

EMSAs. Electrophoretic mobility shift assays (EMSAs) were performed as described previously (28).

In vitro transcription and RNA analysis. (i) With constructs containing the USF binding domain. In vitro transcription and detection of RNA products were done as described previously (28). Because the results were clearer, we used a template containing two USF sites upstream of the HIV core promoter (−41 to +58) rather than the wild-type (wt) ML promoter in front of a G-less cassette as a reporter. As previously shown (28), the two constructs responded to USF in the same way. To distinguish reporter and reference messages in the same reactions, a 3′ truncated version of the HIV promoter (−163 to +22) was ligated to the identical downstream chloramphenicol acetyltransferase sequences as in the reference template, creating a 36-nucleotide-shorter primer extension product.

(ii) With Gal4 fusion proteins. In vitro transcription with Gal4/USF mutants was performed by using plasmid G_5HMC_2AT (containing five Gal4 sites upstream of the ML promoter) as a reporter (100 ng per assay) and the wt ML promoter (−404 to +9) as a reference (40 ng per assay). The two promoters were connected to G-less cassettes of 380 and 240 bp, respectively. Transcription was carried out at 30°C in transcription buffer (100 μM 3′-*O*-methyl-GTP, 25 μM CTP, 500 μM ATP, 500 μM UTP, 7.5 mM $MgCl_2$, 5 mM dithiothreitol, 25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 8.4]) containing 10 μCi of [^{32}P]CTP, 0.4 U of RNase T_1, and 100 μg of HeLa nuclear extract per 30-μl reaction. The purified protein diluted in BC100 was added in 10-, 20-, and 50-fold molar excess over the Gal4 binding sites. Identical results were obtained within this range, indicating that optimal conditions were used. The 30-μl reactions were stopped by addition of 400 μl of stop solution (7 M urea, 10 mM EDTA, 0.5% SDS, 100 mM LiCl, 100 μg of tRNA per ml, 300 mM sodium acetate, 10 mM Tris-Cl [pH 7.8]). Nucleic acids were precipitated by addition of 1 volume of isopropanol. The pellets were washed with 70% ethanol, dried, and resuspended in formamide sample buffer for analysis on a 6% sequencing gel.

RESULTS

Expression of USF deletion mutants. To localize the activation domain(s) of human rUSF^{43}, 14 deletion mutants were constructed (see Materials and Methods) and expressed in bacteria from an IPTG-inducible T7 promoter (34) (Fig. 1A). All mutants contained the (B-HLH) region, including the two cysteines involved in oligomerization, as well as the C-terminal LZ. One set of mutants (Δ1-15 to Δ1-196) contained N-terminal truncations, while a second set (Δ105-130 to Δ26-130) had internal deletions proceeding from the central divergent region toward the N terminus. The central divergent region is the only part of USF^{43} that differs significantly between *Homo sapiens* and *Xenopus* spp. (15). Most mutant proteins were induced well in the expression system of Studier et al. (34), although some of them required several purification steps in addition to the standard ammonium sulfate precipitation procedure (27). Thus, all proteins were purified further on heparin-Sepharose and DEAE 5 PW HPLC columns. The resulting protein preparations were sufficiently pure for all subsequent analyses (Fig. 1B). Despite the overall faster migration of proteins with internal deletions, the changes in apparent molecular weights of the mutant proteins were consistent with the number of deleted residues.

DNA binding properties and transcriptional activity of USF^{43} mutants. The purified USF deletion mutants were tested first for their DNA binding activities. Protein amounts were standardized by reference to SDS-gels stained with Coomassie brilliant blue. Titration in EMSAs revealed similar binding affinities for all mutants on a 39-bp probe containing the USF site from the ML promoter (data not shown). Under saturating conditions (>50-fold molar excess), as used in our transcription assay, the probe (shown without protein in Fig. 2, lane probe) was quantitatively shifted both by full-length rUSF^{43} (lane wt) and by all analyzed B-HLH-LZ-containing USF mutants (Fig. 2). Changes in the mobilities of the shifted complexes corresponded to the different sizes of the mutant proteins.

Transcriptional activation by USF^{43} mutants was analyzed in nuclear extracts immunodepleted of natural USF (28), using both a USF-inducible template (two USF sites upstream of the HIV core promoter; −41 to +58) as a reporter and a USF-independent template (HIV promoter; −163 to +22) as an internal standard (14). For technical reasons (see Materials and Methods), this combination of templates was preferred to templates derived from the ML promoter. However, the reporter template was previously shown to respond to USF in the same way as a template containing the wt ML promoter (28). Specific transcription products were analyzed by primer extension and quantified with a Phosphorimager (Molecular Dynamics). Nuclear extracts immunodepleted of natural USF showed only basal transcriptional activity on both the reporter and reference templates (Fig. 3, lane Depleted NE). Addition of recombinant USF^{43} selectively activated transcription from the reporter template with USF sites, elevating the signal to levels indistinguishable from that observed in the original undepleted nuclear extract (Fig. 3, lane wt) (27). In this assay system, full-length USF^{43} reproducibly stimulated transcription about fivefold. Removal of the 15 N-terminal amino acids had no effect on transcriptional activity. However, deletion of an additional 24 amino acids (Δ1-39) reduced transcriptional activity by 50% compared with the activity of the full-length protein (lane wt). Further deletions reduced transcriptional activation continuously but less dramatically. Transcription with mutants Δ1-156, Δ1-175, and Δ1-196 did not exceed basal levels. Thus, the potential for transcriptional activation was completely lost after removal of the 156 N-terminal amino acids. Two domains, one apparently small and well defined between residues 15 and 59 and a broader one between residues 93 and 156, contributed to transcriptional activation.

An internal deletion removing residues 105 to 130 reduced transcriptional activation by 50% compared with the activity of full-length USF^{43} (Fig. 3). This deletion overlapped the second activation domain indicated by the N-terminal deletions (between residues 93 and 156). It is noteworthy that the deleted residues between positions 105 and 130 were restricted to a region that showed considerable divergence between human and *Xenopus* proteins. Further deletions extending toward the N terminus had little effect on transcriptional activation. However, transcription was reduced almost to basal levels when residues between positions 58 and 26 were included in the internal deletions. This area

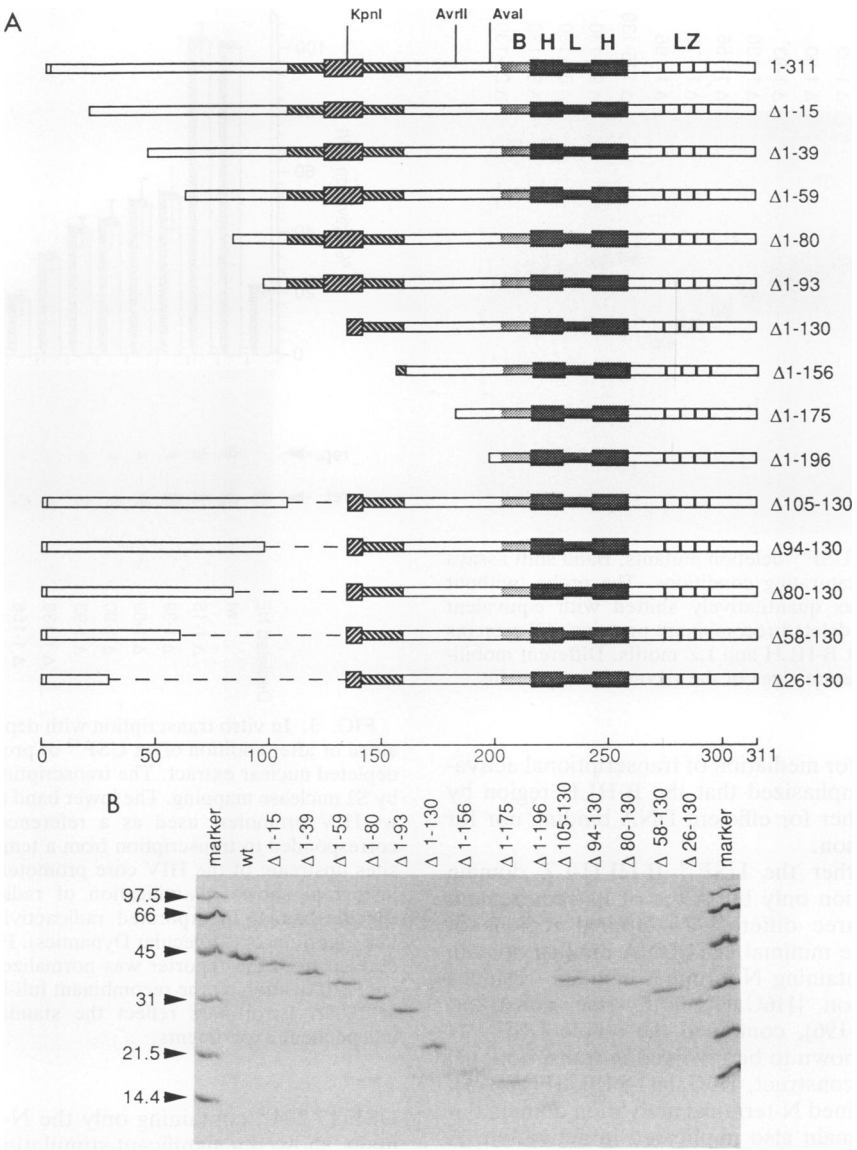

FIG. 1. Bacterially expressed USF[43] deletion mutants. (A) Schematic representation of bacterially expressed wt (1-311) and mutant proteins (indicated with Δ in front of numbers corresponding to the deleted amino acids). The B-HLH and LZ regions are indicated. The *Kpn*I, *Avr*II, and *Ava*I restriction sites in the corresponding cDNAs were used for construction of the deletion mutants (see Materials and Methods). (B) SDS-gel of purified bacterially expressed USF[43] wt and mutant proteins stained with Coomassie brilliant blue. Phosphorylase *b* (97.5 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), and lysozyme (14.4 kDa) were used as molecular size markers.

overlapped the more N-terminal activation domain indicated by the set of N-terminal deletions (compare with the strong effect by removal of residues 16 to 39 in addition to residues 1 to 15). In summary, a domain close to the N terminus of human USF[43] (between positions 16 and 59, with a core between positions 26 and 39) contributed strongly to transcriptional activation. A second domain spanning the region from residues 93 through 156 (with its core between residues 105 and 130) also contributed significantly to transcriptional activation. Full activation clearly required the contribution of both domains.

Transcriptional activation by USF[43] depends on DNA binding but not on the USF B-HLH-LZ domain. Since all mutants tested so far contained the wt B-HLH-LZ USF C terminus, we next addressed the role of these DNA binding and oligomerization domains in transcriptional activation. In a first set of experiments, we deleted the LZ region by introducing a stop codon immediately 3′ terminal of the B-HLH region (following residue 261). The resulting protein (rUHLH) containing an intact B-HLH region showed a markedly reduced ability to bind DNA. In the comparative analysis shown in Fig. 4A, the amount of LZ-deficient USF[43] (16 ng) required for a significant signal by EMSA was between 16 and 48 times the level of intact USF[43] (0.25 to 1 ng) required for a comparable level of binding. In transcription assays, this binding-deficient mutant was inactive (Fig. 4B, UHLH), whereas USF[43] clearly stimulated transcription under the same conditions. Thus, while the DNA binding domain of USF[43] did not activate transcription by itself (Fig. 3, mutant Δ1-196), efficient DNA binding appears to be an

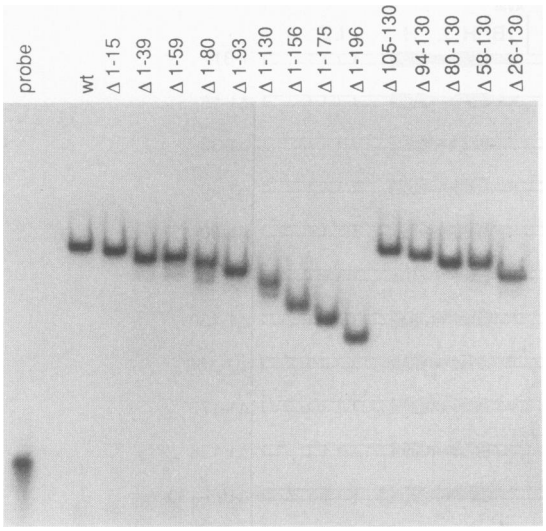

FIG. 2. EMSAs with USF[43] deletion mutants. Band shift assays were performed under saturating conditions. The probe (without protein; lane probe) was quantitatively shifted with equivalent amounts of USF[43] wt or deleted recombinant proteins. All proteins used contained the intact B-HLH and LZ motifs. Different mobilities corresponded to different sizes of the recombinant proteins.

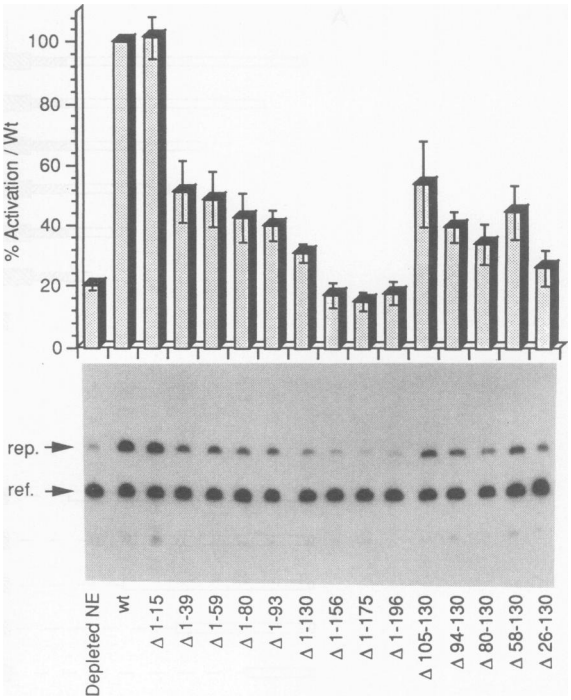

FIG. 3. In vitro transcription with depleted nuclear extract (NE) alone or after addition of wt USF[43] or proteins carrying deletions to depleted nuclear extract. The transcription products were analyzed by S1 nuclease mapping. The lower band (ref.) was derived from the wt HIV promoter, used as a reference; the upper band (rep.) corresponded to transcription from a template containing two USF sites upstream of the HIV core promoter, used as a reporter. The histogram shows quantification of radioactivity incorporated in specific bands. Incorporated radioactivity was measured with a Phosphorimager (Molecular Dynamics). For each column, the signal derived from the reporter was normalized to the respective reference. Activation by the recombinant full-length rUSF[43] (wt) was set at 100%. Error bars reflect the standard deviations from four independent experiments.

absolute requirement for mediation of transcriptional activation. These results emphasized that the B-HLH region by itself is sufficient neither for efficient DNA binding nor for transcriptional activation.

To determine whether the USF[43] B-HLH-LZ domain contributed to activation only by virtue of its requirement for DNA binding, three different N-terminal regions of USF were fused to the minimal Gal4 DNA binding domain (residues 1 to 94) containing N-terminal histidine residues for affinity purification [H6Gal4(1-94)]. One construct, H6Gal4(1-94)/USF(17-196), contained the whole USF[43] N terminus previously shown to be involved in transcriptional activation. A second construct, H6Gal4(1-94)/USF(17-104), contained the well-defined N-terminal activation domain but lacked the central domain also implicated in activation. A third construct, H6Gal4(1-94)/USF(131-196), lacked the N-terminal activation domain as well as most of the second, central activation domain. The recombinant proteins were expressed in a pDS-derived bacterial expression system (see Materials and Methods) and batch purified to near homogeneity on nickel resin (Fig. 5A). B-HLH-LZ-independent transcriptional activation was tested on a template containing five Gal4 sites upstream of the HIV TATA box and the ML initiator in front of a G-less cassette (380 bp). A template containing the ML promoter (positions −404 to +9) connected to a shorter (240 bp) G-less cassette was used as a reference. In unfractionated nuclear extracts, the basal activity of the reporter template (upper band) was stimulated to various degrees by the various H6Gal4(1-94) derivatives, whereas the activity of the reference template (lower band) remained unaltered (Fig. 5B). In agreement with the previously documented presence of a cryptic activation domain in Gal4(1-94), this protein activated transcription twofold in this assay. However, significantly stronger effects were seen with derivatives containing the USF[43] activation domains described above. H6Gal4(1-94)/USF(17-196) showed the strongest effect, increasing transcriptional activation at least threefold compared with H6Gal4(1-94) alone. H6Gal4(1-94)/USF(17-104), containing only the N-terminal activation domain, showed a significant stimulation relative to H6Gal4(1-94) alone but a reduced effect relative to the derivative with both domains. Finally, transcription with the fusion protein H6Gal4(1-94)/USF(131-196) lacking all or most of both domains did not significantly exceed levels induced by H6Gal4(1-94) alone. In summary, these results show that N-terminal USF[43] domains have an ability to activate transcription independently of the major domain (B-HLH-LZ) essential for DNA binding and that both N-terminal domains involved in transcriptional activation are required for optimal activity.

DISCUSSION

Using deletion mutants and in vitro transcription assays, we have characterized the domains of USF[43] involved in mediating transcriptional activation. This approach has been chosen in favor of transfection assays because of the high natural background of endogenous USF in cell lines. It also has taken advantage of the ability to selectively remove USF[43] from nuclear extracts by immunodepletion and the ability of recombinant USF[43] to fully restore activation of USF-responsive reporters. Fourteen N-terminal and internal deletion mutants of USF[43], one C-terminal deletion, and

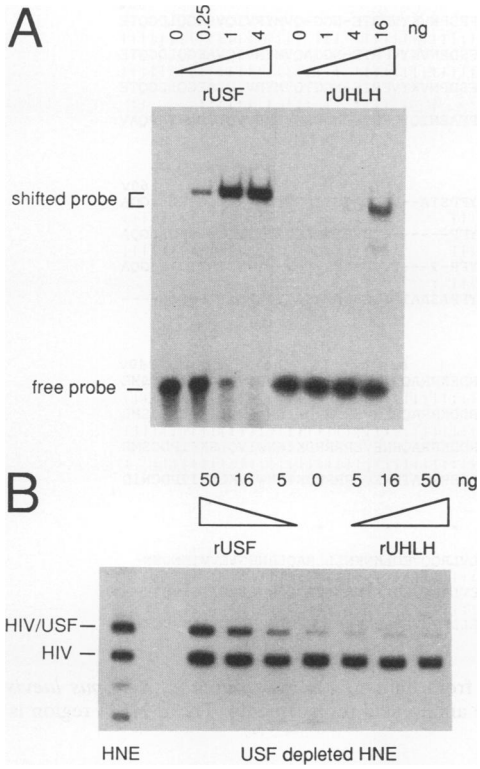

FIG. 4. DNA binding and transcriptional activation by full-length USF[43] and a C-terminally truncated USF[43] mutant protein. (A) EMSAs with wt USF[43] (rUSF) and C-terminally truncated USF lacking the LZ (rUHLH). Binding activity was titrated by using 0 to 4 ng (wt USF) or 0 to 16 ng (rUHLH) of highly purified protein. The higher mobility of the shifted band with rUHLH is due to the truncation of the protein. (B) Titration of transcriptional activation by wt rUSF and rUHLH. Recombinant protein (5, 16, or 50 ng) was added to the transcription reaction mixtures. Transcription products were detected by primer extension. The lower band corresponds to HIV wt template (−163 to +22), used as a control. The upper band is derived from the reporter plasmid containing two USF sites upstream of the HIV core promoter (−41 to +22). HNE, HeLa nuclear extract.

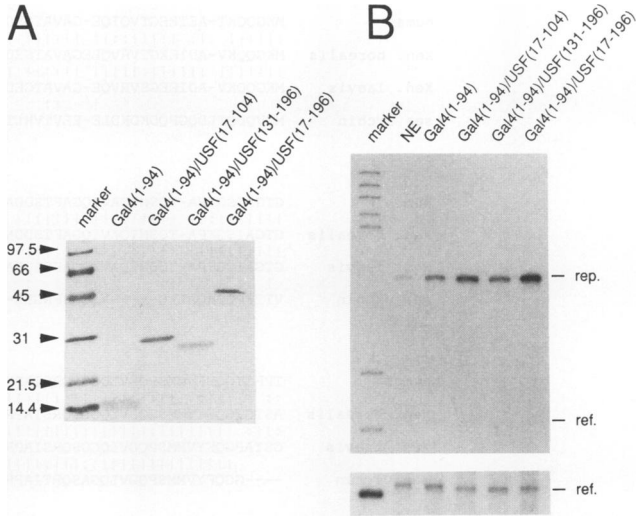

FIG. 5. Transcriptional activation by Gal4/USF fusion proteins. (A) SDS-gel electrophoresis with affinity-purified Gal4(1-94) and Gal4(1-94)/USF fusion proteins. Molecular sizes were those used for Fig. 1B. (B) Gal4 binding-dependent transcriptional activation on a template containing five Gal4 sites upstream of a G-less cassette of 380 bp. The wt ML promoter upstream of a truncated G-less cassette (240 bp) was used as a reference (ref.). The ML-derived signal was kept low by reducing the amount of template. The lower panel shows a longer exposure of the 240-nucleotide band. End-labeled fragments from a *Hae*III digest of pBR322 (Boehringer Mannheim) served as a size marker. NE, nuclear extract; rep., reporter gene.

three Gal4/USF fusion proteins containing different parts of USF[43] fused to H6Gal4(1-94) were analyzed on USF- or Gal4-responsive templates. Analysis of the recombinant USF[43] derivatives led to the identification of two USF[43] activation domains, one between residues 15 and 59 and another between residues 93 and 156. Full activation depended on the presence of both domains. Involvement of several distinct intramolecular domains in transcriptional activation has been shown for many transcriptional activators. An interesting parallel to USF[43] is found in C/EBP. The N-terminal part of C/EBP can be subdivided into two distinct transactivation domains interrupted by an attenuator region (24). Although we have no evidence for an attenuator sequence, the present results demonstrate a bipartite activation domain in USF[43]. At least two other members of the HLH family of transcriptional activators contain an activation domain close to the N terminus. In MyoD, an activation domain maps within the amino-terminal 53 residues (37). In the case of TFE3, a strong N-terminal activation domain contains a clear amphipathic helix (2) described as an important class of eukaryotic transcription activation domains (11, 29).

A variety of characteristics have been attributed to transcriptional activation domains (for a review, see reference 21). The USF[43] sequence, however, lacks any regions obviously rich in glutamine, proline, or acidic residues. In addition, the overall abundance of serine/threonine residues N terminal of the B-HLH domain (22% of the first 199 residues) does not seem to be concentrated in any particular subregion. However, it cannot be excluded that serines and threonines from different parts of the N-terminal domain are clustered as a result of specific protein folding. Upon phosphorylation, these residues could form an acidic activation domain (for a review, see reference 29). Although hypothetical, the creation of an acidic activation domain by phosphorylation was postulated for Oct-2 (36). The recombinant proteins analyzed here were expressed in bacteria and presumably underphosphorylated, but phosphorylation might have occurred upon addition to the nuclear extract during the in vitro transcription assay. Apart from this highly speculative mechanism and the additional notion that the N-terminal activation domain (between residues 15 and 59) contains the major hydrophobic domain within USF[43] (between residues 29 and 41), we cannot attribute any specific chemical characteristics to the USF domains shown to be involved in transcriptional activation.

Surprisingly, the domains responsible for transcriptional activation by USF[43] are not restricted to highly conserved sequences among different species. Deletions of residues 105 to 130 or 130 to 156 significantly reduced the potential for transcriptional activation (Fig. 3), although these residues included the only part of USF[43] (residues 108 to 152) poorly conserved between human and *Xenopus* proteins (15) (Fig. 6). Moreover, transcriptional activation was strongly reduced by the deletion of residues 16 to 39, a region that is

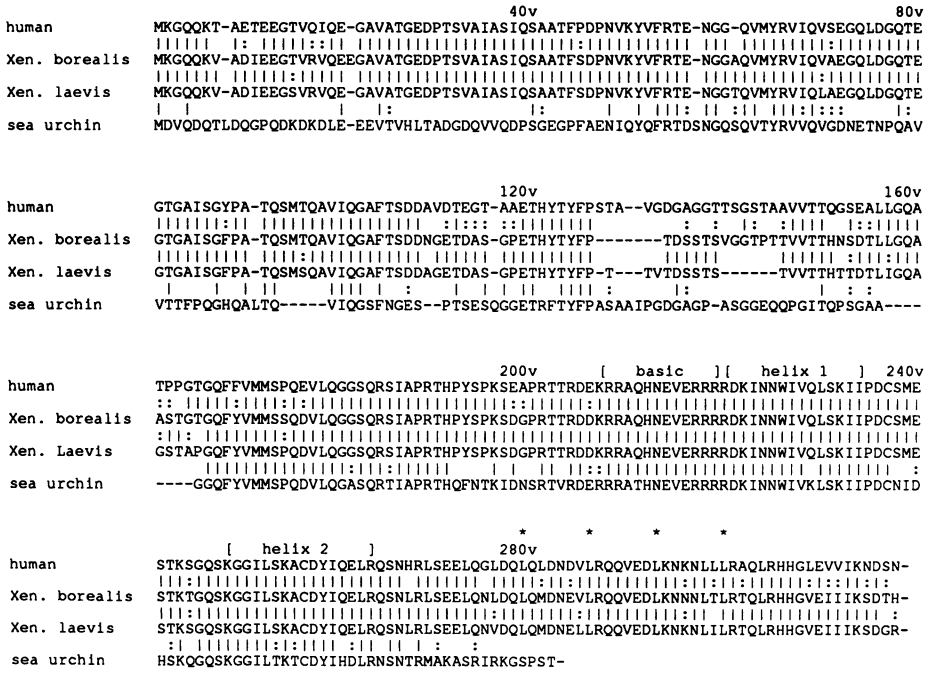

FIG. 6. Amino acid sequence comparison between USF sequences derived from human, *Xenopus borealis*, *Xenopus laevis*, and sea urchin. Vertical bars signify amino acid identities; colons represent conservative amino acid replacements. The B-HLH region is indicated by brackets. The leucines forming the C-terminal LZ are marked with asterisks.

conserved between human and *Xenopus* proteins but not in sea urchin USF (16) (Fig. 6). In this context, it should be noted that the sea urchin homolog to USF[43] shows high sequence conservation only in the midportion of the protein that includes the B-HLH region. Most significantly, it lacks the C-terminal LZ, a region that is essential for efficient DNA binding and consequent transcriptional activation by human USF[43] (Fig. 4). Sea urchin USF still binds specifically to DNA (16). However, no data for its transcriptional activation potential have yet been shown. The structural differences between various species could indicate distinct mechanisms in USF function. Assuming that *Xenopus* and sea urchin homologs of USF[43] also mediate transcriptional activation by analogous means, the functional importance of divergent regions could point to less stringent selection criteria for transcriptional activation domains than for DNA binding or oligomerization domains. Some of the additional highly conserved regions could be responsible for the nuclear localization or functional interactions with other factors. The nuclear localization signal in c-*myc*, for instance, lies just N terminal of the B-HLH motif (for a review, see reference 19). The corresponding part of USF[43] is highly conserved among all four characterized species as well as in a new USF-related protein that interacts with c-*fos* (3). Our results do not provide any evidence for the involvement of residues in the B-HLH-LZ in transcriptional activation per se but rather suggest an indirect involvement via DNA binding.

It is possible that USF[43] activation domains are not directly responsible for transcriptional activation but exert their influence via interactions with other proteins. Indeed, in contrast to other activators such as SP1, USF[43] does not activate transcription in highly purified systems (20). Obvious candidates for functional partners are USF[44], another component (together with USF[43]) in natural USF (32), and TFII-I (30). Involvement of USF[44] in the activation observed here appears to be highly unlikely, since immunodepletion of nuclear extracts with antibodies against USF[43] also leads to depletion of USF[44] below levels detectable by Western immunoblot blot or EMSA (28). It seems possible that the endogenous TFII-I is able to contribute to the transcriptional activation by USF[43]. Rate-limiting functional interactions of USF[43] with other, possibly less abundant factors under certain conditions could explain a tight regulation by heterodimeric complexes despite the general abundance of USF. Such interactions could be responsible, for example, for the regulation of the ML promoter during adenovirus infection even though USF binding activities remain constant (18). The factors necessary and sufficient for transcriptional activation by USF[43] as well as additional factors that might be able to functionally interact with USF[43] under certain conditions remain to be determined.

ACKNOWLEDGMENTS

B.J.K. and P.P. contributed equally to this work.

We thank C. Rosen for generously providing the His-tag Gal4 expression system, H. Kato for providing the 2xUSF-HIV template, and R. Bernstein for providing the G_5HMC_2AT template. We also thank members of the Roeder laboratory, in particular E. Martinez, A. Roy, J.-B. Yoon, and C. Parada, for support and encouragement and for discussions and help during preparation of the manuscript.

This work was supported in part by Public Health Service grants CA 42567 and AI 27397 to R.G.R. and by grants from the Pew Charitable Trusts to The Rockefeller University. B.J.K. is a Lita Annenberg Hazen postdoctoral fellow at The Rockefeller University. P.P. is supported by a long-term EMBO postdoctoral fellowship.

REFERENCES

1. **Beckmann, H., and T. Kadesch.** 1991. The leucine zipper of TFE3 dictates helix-loop-helix dimerization specificity. Genes Dev. 5:1057–1066.

2. **Beckmann, H., L.-K. Su, and T. Kadesch.** 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev. **4**:167–179.
3. **Blanar, M. A., and W. J. Rutter.** 1992. Interaction cloning: FIP encodes a helix-loop-helix-zipper protein that interacts with cfos. Science **256**:1014–1018.
4. **Carr, C. S., and P. A. Sharp.** 1990. A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. Mol. Cell. Biol. **10**:4384–4388.
5. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43**:439–448.
6. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1987. The major late transcription factor binds to and activates the mouse metallothionein I promoter. Genes Dev. **1**:973–980.
7. **Certa, U., W. Bannwarth, D. Stüber, R. Gentz, M. Lanzer, S. LeGrice, F. Guillot, I. Wendler, G. Hunsmann, H. Bujard, and J. Mous.** 1986. Subregions of a conserved part of the HIV gp41 transmembrane protein are differentially recognized by antibodies of infected individuals. EMBO J. **5**:3051–3056.
8. **Chodosh, L. A., R. W. Carthew, J. G. Morgan, G. R. Crabtree, and P. A. Sharp.** 1987. The adenovirus major late transcription factor activates the rat γ-fibrinogen promoter. Science **238**:684–688.
9. **Chodosh, L. A., R. W. Carthew, and P. A. Sharp.** 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. **6**:4723–4733.
10. **Gentz, R., C.-H. Chen, and C. A. Rosen.** 1989. Bioassay for trans-activation using purified human immunodeficiency virus *tat*-encoded protein: trans-activation requires mRNA synthesis. Proc. Natl. Acad. Sci. USA **86**:821–824.
11. **Giniger, E., and M. Ptashne.** 1987. Transcription in yeast activated by a putative amphipathic α helix linked to DNA-binding unit. Nature (London) **330**:670–672.
12. **Gregor, P. D., M. Sawadogo, and R. G. Roeder.** 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev. **4**:1730–1740.
13. **Hu, Y. F., B. Lüscher, A. Admon, N. Mermod, and R. Tjian.** 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. **4**:1741–1752.
14. **Kato, H., H. Sumimoto, P. Pognonec, C. H. Chen, C. A. Rosen, and R. G. Roeder.** 1992. HIV-1 Tat acts as a processivity factor in vitro in conjunction with cellular elongation factors. Genes Dev. **6**:665–666.
15. **Kaulen, H., P. Pognonec, P. D. Gregor, and R. G. Roeder.** 1991. The *Xenopus* B1 factor is closely related to the mammalian activator USF and is implicated in the developmental regulation of TFIIIA gene expression. Mol. Cell. Biol. **11**:412–424.
16. **Kozlowski, M. T., L. Gan, J. M. Venuti, M. Sawadogo, and W. H. Klein.** 1991. Sea urchin USF: a helix-loop-helix protein active in embryonic ectoderm cells. Dev. Biol. **148**:625–630.
17. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
18. **Leong, K. L., L. Brunet, and A. J. Berk.** 1988. Factors responsible for the higher transcriptional activity of extracts of adenovirus-infected cells fractionate with the TATA box transcription factor. Mol. Cell. Biol. **8**:1765–1774.
19. **Lüscher, B., and R. N. Eisenmann.** 1990. New light on myc and myb. Part I. Myc. Genes Dev. **4**:2025–2035.
20. **Meisterernst, M., and R. G. Roeder.** Unpublished data.
21. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
22. **Miyamoto, N. G., V. Moncollin, J. M. Egly, and P. Chambon.** 1985. Specific interaction between a transcription factor and the upstream element of the adenovirus-2 major late promoter. EMBO J. **4**:3563–3570.
23. **Moncollin, V., N. G. Miyamoto, X. M. Zheng, and J. M. Egly.** 1986. Purification of a factor specific for the upstream element of the adenovirus-2 major late promoter. EMBO J. **5**:2577–2584.
24. **Pei, D., and C. Shih.** 1991. An "attenuator domain" is sandwiched by two distinct transactivation domains in the transcription factor C/EBP. Mol. Cell. Biol. **11**:1480–1487.
25. **Peritz, L. N., E. J. B. Fodor, D. W. Silversides, P. A. Cattini, J. D. Baxter, and N. L. Eberhardt.** 1988. The human growth hormone gene contains both positive and negative control elements. J. Biol. Chem. **263**:5005–5007.
26. **Pognonec, P., H. Kato, and R. G. Roeder.** The helix-loop-helix/leucine repeat transcription factor USF can be functionally regulated in a redox dependent manner. J. Biol. Chem., in press.
27. **Pognonec, P., H. Kato, H. Sumimoto, M. Kretzschmar, and R. G. Roeder.** 1991. A quick procedure for purification of functional recombinant proteins over-expressed in E. coli. Nucleic Acids Res. **19**:6650.
28. **Pognonec, P., and R. G. Roeder.** 1991. Recombinant 43-kDa USF binds to DNA and activates transcription in a manner indistinguishable from that of natural 43/44-kDa USF. Mol. Cell. Biol. **11**:5125–5136.
29. **Ptashne, M.** 1988. How eukaryotic transcriptional activators work. Nature (London) **335**:683–689.
30. **Roy, A. L., M. Meisterernst, P. Pognonec, and R. G. Roeder.** 1991. Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature (London) **354**:245–248.
31. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
32. **Sawadogo, M., M. V. Van Dyke, P. D. Gregor, and R. G. Roeder.** 1988. Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei. J. Biol. Chem. **263**:11985–11993.
33. **Scotto, K. W., H. Kaulen, and R. G. Roeder.** 1989. Positive and negative regulation of the gene for transcription factor IIIA in *Xenopus laevis* oocytes. Genes Dev. **3**:651–662.
34. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185**:60–89.
35. **Stueber, D., I. Ibrahimi, D. Cutler, B. Dobberstein, and H. Bujard.** 1984. A novel in vitro transcription-translation system: accurate and efficient synthesis of single proteins from cloned DNA sequences. EMBO J. **13**:3143–3148.
36. **Tanaka, M., and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. Cell **60**:375–386.
37. **Weintraub, H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott.** 1991. Muscle-specific transcriptional activation by MyoD. Genes Dev. **5**:1377–1386.