

# Cloning of an Inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1

Ananda L.Roy<sup>1,2</sup>, Hong Du<sup>1,3</sup>,  
Polly D.Gregor<sup>1,4</sup>, Carl D.Novina<sup>1,2</sup>,  
Ernest Martinez<sup>1</sup> and Robert G.Roeder<sup>1,5</sup>

<sup>1</sup>Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021,

<sup>2</sup>Department of Pathology and Program in Immunology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111, <sup>3</sup>Division and Program of Human Genetics, Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati, Cincinnati, OH 45229 and <sup>4</sup>Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA

<sup>5</sup>Corresponding author

A.L.Roy and H.Du contributed equally to this work

**The transcription factor TFII-I has been shown to bind independently to two distinct promoter elements, a pyrimidine-rich initiator (Inr) and a recognition site (E-box) for upstream stimulatory factor 1 (USF1), and to stimulate USF1 binding to both of these sites. Here we describe the isolation of a cDNA encoding TFII-I and demonstrate that the corresponding 120 kDa polypeptide, when expressed ectopically, is capable of binding to both Inr and E-box elements. The primary structure of TFII-I reveals novel features that include six directly repeated 90 residue motifs that each possess a potential helix-loop/span-helix homology. These unique structural features suggest that TFII-I may have the capacity for multiple protein-protein and, potentially, multiple protein-DNA interactions. Consistent with this hypothesis and with previous *in vitro* studies, we further demonstrate that ectopic TFII-I and USF1 can act synergistically, and in some cases independently, to activate transcription *in vivo* through both Inr and the E-box elements of the adenovirus major late promoter. We also describe domains of USF1 that are necessary for its independent and synergistic activation functions.**

**Keywords:** adenovirus major late promoter/E-box/Inr/transcription factor TFII-I/upstream activation factor 1

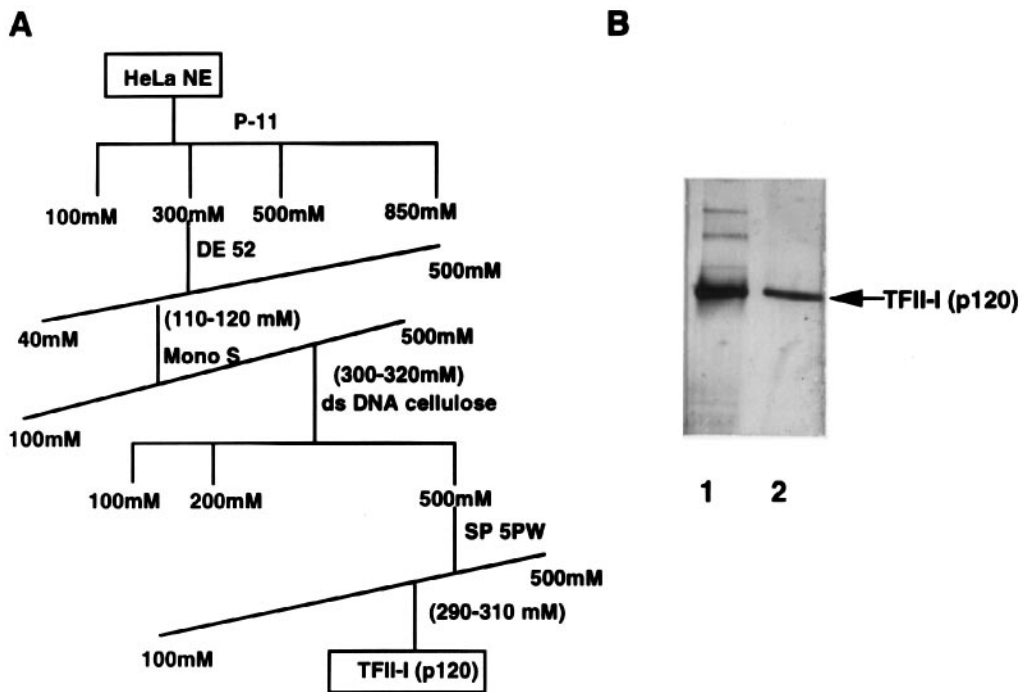
## Introduction

Transcription of eukaryotic protein-coding genes is initiated by interactions of RNA polymerase II and general initiation factors at common core promoter elements, and is regulated by various gene-specific activators that act through adjacent or distal regulatory elements (Roeder 1991, 1996; Conaway and Conaway, 1993; Zawel and Reinberg, 1995; Orphanides *et al.*, 1996). Therefore, communication between the activation and basal (core promoter) components of eukaryotic transcription is critical for appropriate gene expression. In metazoans, the most common core promoter elements, which can act

independently or in concert to determine the transcription start site, are the TATA box near position -30 and a pyrimidine-rich initiator (Inr) element (consensus YYA<sub>+1</sub>NT/AYY) located near the start site (Breathnach and Chambon, 1981; Smale and Baltimore, 1989; Javahery *et al.*, 1994).

With respect to core promoter functions, the minimal factor requirement and corresponding pre-initiation complex (PIC) assembly pathway are best understood for TATA-directed basal transcription, in which case TATA recognition by the TATA-binding protein (TBP) component of TFIID is sufficient to nucleate the assembly of other general initiation factors and RNA polymerase II into a functional complex (Roeder, 1996). The pyrimidine-rich Inr-directed basal transcription is more complicated and less well understood, but requires several factors, including both the TBP-associated factor (TAF) subunits of TFIID and other novel factors, that are not required for TATA-directed transcription (Martinez *et al.*, 1994; Roeder, 1996; Smale, 1997). Factors which have been demonstrated or inferred to recognize the Inr and to nucleate PIC assembly include the TAF components of TFIID, RNA polymerase II and novel Inr-binding proteins (reviewed in Smale, 1997). Consistent with the latter possibility, several factors have been reported to bind at or near Inr elements (Novina and Roy, 1996; Smale, 1997) and, in some cases, shown to facilitate core promoter functions *in vitro* (Roy *et al.*, 1991, 1993a; Seto *et al.*, 1991; Usheva and Shenk, 1994). This multiplicity of Inr-binding proteins could reflect diversity in core promoter elements, especially in view of the loose consensus for such elements (Kaufman *et al.*, 1996). Alternatively, as suggested (Wiley *et al.*, 1992; Kaufman and Smale, 1994), these observations could also reflect juxtaposition or overlap of binding sites for various regulatory factors and Inr sites that could be recognized by a universal (but still unidentified) factor. In any case, what is needed to settle the issue unequivocally is identification and characterization of the protein factors directly involved in Inr function.

TFII-I was identified originally as a factor that could bind to Inr elements and stimulate transcription from the potent TATA- and Inr-containing adenovirus major late (AdML) promoter in a system reconstituted with partially purified components (Roy *et al.*, 1991, 1993a). Somewhat surprisingly, TFII-I was also found to bind to a distinct upstream element (E-box) on the AdML promoter that originally was identified as a recognition site for the transcriptional activator USF, a member of the basic helix-loop-helix-leucine zipper (bHLH-LZ) family of proteins (reviewed in Murre and Baltimore, 1992) that activates the AdML promoter both *in vitro* and *in vivo* (Pognonec and Roeder, 1991; Du *et al.*, 1993; Luo and Sawadogo, 1996). Similarly, USF1 was also shown to bind not only to the E-box but also to the Inr (Roy *et al.*, 1991; Du



**Fig. 1.** Purification of TFII-I. (A) Purification scheme of native TFII-I. (B) Silver-stained dsDNA cellulose (lane 1) and HPLC-purified (lane 2) native TFII-I subsequent to SDS-PAGE.

*et al.*, 1993). Consistent with these observations, as well as synergistic interactions at both Inr and E-box elements, ectopic expression of USF1 was found to enhance expression of TATA-containing promoters either through an adjacent Inr element (AdML promoter) or through upstream E-boxes (E1b promoter) (Du *et al.*, 1993). Although it is not yet clear whether USF1 is unique with respect to its apparent dual function through two distinct promoter elements, and whether these functions might be linked in some promoters, these observations suggested novel mechanisms of gene regulation and the possible involvement of TFII-I as a co-regulator that can integrate regulatory responses of USF1 to the basal machinery. The involvement of such co-regulators may also help explain the differential functions of distinct bHLH-LZ proteins through common E-box elements in different promoters (Weintraub *et al.*, 1994; Molkenin *et al.*, 1995).

As part of our investigation of these questions, we now report the purification of native TFII-I and the cloning of a cognate cDNA whose ectopically expressed product, like its native counterpart, exhibits specific Inr- and E-box-dependent binding. Consistent with our model, the ectopically expressed TFII-I markedly enhances both USF1 binding *in vitro* and, most importantly, the *in vivo* function of ectopic USF1 through both Inr and E-box elements in the AdML promoter. Taken together, these results indicate that TFII-I may serve as a novel co-regulator for USF1 in addition to, or in conjunction with, its potential role as an Inr-binding basal transcription factor.

## Results

### Purification of TFII-I

Using an electrophoretic mobility shift assay (EMSA) to monitor site-specific binding to the AdML promoter Inr element (Roy *et al.*, 1991), TFII-I was purified according

to the scheme shown in Figure 1A and detailed in Materials and methods. The TFII-I activity eluted predominantly with a 120 kDa polypeptide at the dsDNA cellulose step (Figure 1B, lane 1) and exclusively with this polypeptide at the final HPLC (SP-5PW) step (Figure 1B, lane 2 and data not shown).

### Primary structure of TFII-I

The purified material from the SP-5PW HPLC column was resolved by SDS-PAGE, transferred to nitrocellulose and subjected to microsequencing. The 120 kDa polypeptide yielded four peptide sequences, indicated by underlining in Figure 2A, three of which were used to design primers for screening a Namalwa (B cell)-derived cDNA library (Scheidereit *et al.*, 1988). Extensive screening yielded a cDNA clone with a 957 amino acid open reading frame (ORF) that was unique (GenBank database) and contained all four peptide sequences derived from microsequencing (Figure 2A). Most strikingly, analysis of the amino acid sequence (Figure 2B) revealed six direct repeats (R1-R6), each 90 amino acids long, suggesting that TFII-I probably arose via gene duplication. The internal or core repeats, R2-R5, are more closely related to each other than to either of the flanking repeats (R1 and R6). The remarkable sequence conservation amongst R2-R5 is highlighted by a region (underlined) that is nearly identical amongst these repeats. Several other interesting structural features also are apparent. First, the presence of a hydrophobic zipper-like region at the N-terminal portion of the protein (indicated by bold amino acids, Figure 2A) suggests a protein interaction domain, although the functional significance of this zipper-like region is not known at present. Moreover, unlike the conventional basic leucine zipper DNA-binding proteins, this region is not flanked by a conserved basic region that could be involved in DNA binding (Ferre-D'Amare *et al.*,

**A**

```

MAQVAMSTLPEVEDEESSESRMV+VTF+FLMSALESMCKELAKSKAEVACIAVY 50
ETDVFVVGTERGRAFVNTRKDFQKDFVKYCV+EEEEKAAEMHKMKSTTQAN 100
RMSVDAVEIETLRKTVEDYFCFCYGKALGKSTVVPVPEYKMLRDQSAVVV 150
QGLPEGVAFKHPENYDLATLKWILENKAGISFI+IKRPFLEPKKHVGGGRVM 200
VTDADRSILSPGGSCGPIKVKTEPTEDSGISLEMAAVTVKEESEDPPDYQ 250
YNIQGS+SHHSE+SEGNEGTEMEVPAEDDDYSPPSKRPKANELPQPPVPEPANA 300
GKRKBRVREFNFEKWNARITDLRBRKQVEELFERKYAQAIKAKGPVTIPYPLFQ 350
SHVEDLYVEGLPEGIPFRPSTYGI+PRLERILLAKERIRFVIKKHELLNS 400
TREDLQDKPASGVKEEWYARITKLRKMVDQLFCKKFAERLGSTEAKAVP 450
YQKFEAHPNDLYVEGLPENIPFRSPSWYGI+PRLEKIIQVGNRIKFVIKRP 500
ELLTHSTTEVTQPRNT+PKEDWNVRITKLRKQVEE+IFNLKFAQALGLTE 550
AVKVPYPVFESNPEFLYVEGLPEGIPFRSPTWFGI+PRLERIVRGSNKIKF 600
VVKKPELVISYLP+PGMA+SKINTKALQSPKRPSPGNSKVP+EIEVTVVEGP 650
NNNNPQTS+AVRTPQTNGSNV+PFKPRGREFSFEAWNAKITDLKQKVENLF 700
NEKCGEALGLKQAVKVPFALFESFPEDFYVEGLPEGV+FRPSTFGIPRL 750
EKILRNKAKIKFI+IKKPEMFETA+IKESTSSKSP+PRKINS+SPNVNTTASGV 800
EDLNI+IQVTIPDDN+ERLSKVEKARQLRE+QVNDLFSRKFGEAIGMGFPVK 850
VPYRKITINPGCVVDGMP+PGVSFKAPSYLEISSMRRILDSAEFIKFTYI 900
RPF+PGLVIN+QLVDQSESEGPVIQESAEP+SQLEVPATEEIKETDGSSQIK 950
QEPDPTW
    
```

**B**

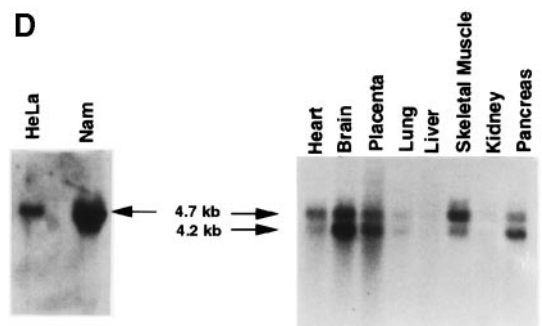
102	191	
MSVDAVEIETLRKTVEDYFCFCYGKALGKSTVVPVPEYKMLRDQSAVVVQGLPEGVAFKHPENYDLATLKWILENKAGISFI <sup>+</sup> IKRPFLEP	R1	310
310	399	
F <sup>BR</sup> EKWNARITDLR <sup>BR</sup> KQVEELFERKYAQAIKAKGPVTIPYPLFQSHVEDLYVEGLPEGIPFRPSTYGI <sup>+</sup> PRLERILLAKERIRFVIKKHELLN	R2	415
415	504	
KEEWYARITKLRKMVDQLFCKKFAEALGSTEAKAVPYQKFEAHPNDLYVEGLPENIPFRSPSWYGI <sup>+</sup> PRLEKIIQVGNRIKFVIKRPPELLT	R3	520
520	609	
KEDWNVRITKLRKQVEE <sup>+</sup> IFNLKFAQALGLTEAVKVPYPVFESNPEFLYVEGLPEGIPFRSPTWFGI <sup>+</sup> PRLERIVRGSNKIKFVVKKPELVI	R4	682
682	771	
FEAWNAKITDLKQKVENLFNEKCGEALGLKQAVKVPFALFESFPEDFYVEGLPEGVFRPSTFGIPRLKILRNKAKIKFI <sup>+</sup> IKKPEMFE	R5	817
817	906	
RLSKVEKARQLRE <sup>+</sup> QVNDLFSRKFGEAIGMGFPVKVPYRKITINPGCVVDGMP <sup>+</sup> PGVSFKAPSYLEISSMRRILDSAEIGIKFTVIRPFPGL	R6	

-----I--LR--V--F--K--A-G---V-VPY-----V-GLPEG--F--P---I--L--I-----I-F--K-P----

**C**

MYC	NELKRSFFALRDQ---KVVILKKATAYIL
USF	DKINNMIVQLSKI---KGGILSKACDYIQ
TFII-I	EKWNARITDLRQ---KHELLNSTREDLQ
Consensus	Φ Ω L Q---K LL L

Helix-Loop-Helix Homology



**Fig. 2.** Amino acid sequence of TFII-I. (A) Primary structure of TFII-I protein indicating the four peptides (underlined) derived from microsequencing. A leucine zipper-like region is indicated by bold amino acids (VLLV). Acidic amino acid regions are indicated by overhead + signs. The putative basic region preceding repeat 2 (R2) is overlined and indicated as BR. A peptide comprised of amino acids 301–321, which included the BR, was employed to generate the anti-peptide antibody. A consensus MAPK site (PRSP) is apparent at amino acids 631–634. Src autophosphorylation sites (EDXDY) are at positions 244–248 and 273–277. Finally, a putative SH3 recognition helix is present at positions 290–297. (B) Arrangement of six direct repeats in TFII-I, starting from position 102 and extending to position 906 with the internal (core) repeats (R2–R5) showing a closer sequence relationship to each other than to the flanking repeats (R1 and R6). In turn, the flanking repeats are more closely related to each other than to the internal repeats. The most highly conserved amino acids are indicated at the bottom. The amino acids in bold represent identity in all six repeats. Amino acids that are conserved in at least five of the repeats are also indicated. The most conserved region within these repeats is indicated by the solid line and is termed the ‘I-repeat’. (C) The putative helix–loop/span–helix homology in TFII-I compared with the helix–loop–helix proteins USF and c-MYC. For the sake of simplicity, only the HLH homology in R2 is shown. Other repeats also have similar homology. There is a greater identity to the USF sequence (indicated by solid lines) than to the c-MYC sequence. (D) Northern blot analysis on poly(A)<sup>+</sup> RNA isolated from HeLa and Namalwa (Nam) shows a predominant 4.7 kb TFII-I RNA (left panel). It is about three times more abundant in Namalwa cells than in HeLa cells. Northern analysis on a multiple tissue blot shows that although the TFII-I RNA is ubiquitously expressed, the levels vary significantly in various tissues (right panel). Moreover, these tissues contain a second (4.2 kb) RNA whose exact relationship to the 4.7 kb RNA is not clear yet.

1993). Second, the region within the N-terminal 90 amino acids, and before the beginning of R1, includes two clusters of four acidic amino acids. A third acidic cluster is also apparent between R1 and R2 (indicated by + signs, Figure 2A). Although the functional significance of these acidic clusters is uncertain at present, they are reminiscent of acidic activation domains present in eukaryotic transcriptional activator proteins (Triezenberg, 1995). Importantly, all of these special structural features in TFII-I lie outside of or between the direct repeats in the 'linker' regions.

Careful analysis of the primary amino acid structure of TFII-I demonstrated a putative HLH-like domain (Figure 2C) within each of the repeats. However, there appears to be only one putative basic region (BR, between amino acids 301 and 321) that, by analogy to known basic-HLH domain proteins, could constitute a DNA-binding domain. In contrast to the conventional HLH domains in which the loop ranges from six to 20 amino acids (reviewed in Ferre-D'Amare *et al.*, 1993), but more like the long loop region in AP-4 (Hu *et al.*, 1990), the loop region in TFII-I is ~70 amino acids. This fact, and the presence of multiple putative HLH motifs, makes TFII-I a unique transcription factor, potentially capable of interacting with a variety of HLH regulators (Roy *et al.*, 1993b; Roy and Roeder, 1994). The presence of multiple HLH-like motifs also raises the possibility that in addition to forming intermolecular heteromeric interactions with other classical HLH proteins, TFII-I may dimerize intramolecularly and thereby display different configurations (e.g. two distinct DNA-binding domains) depending on the particular combination of intramolecular interactions.

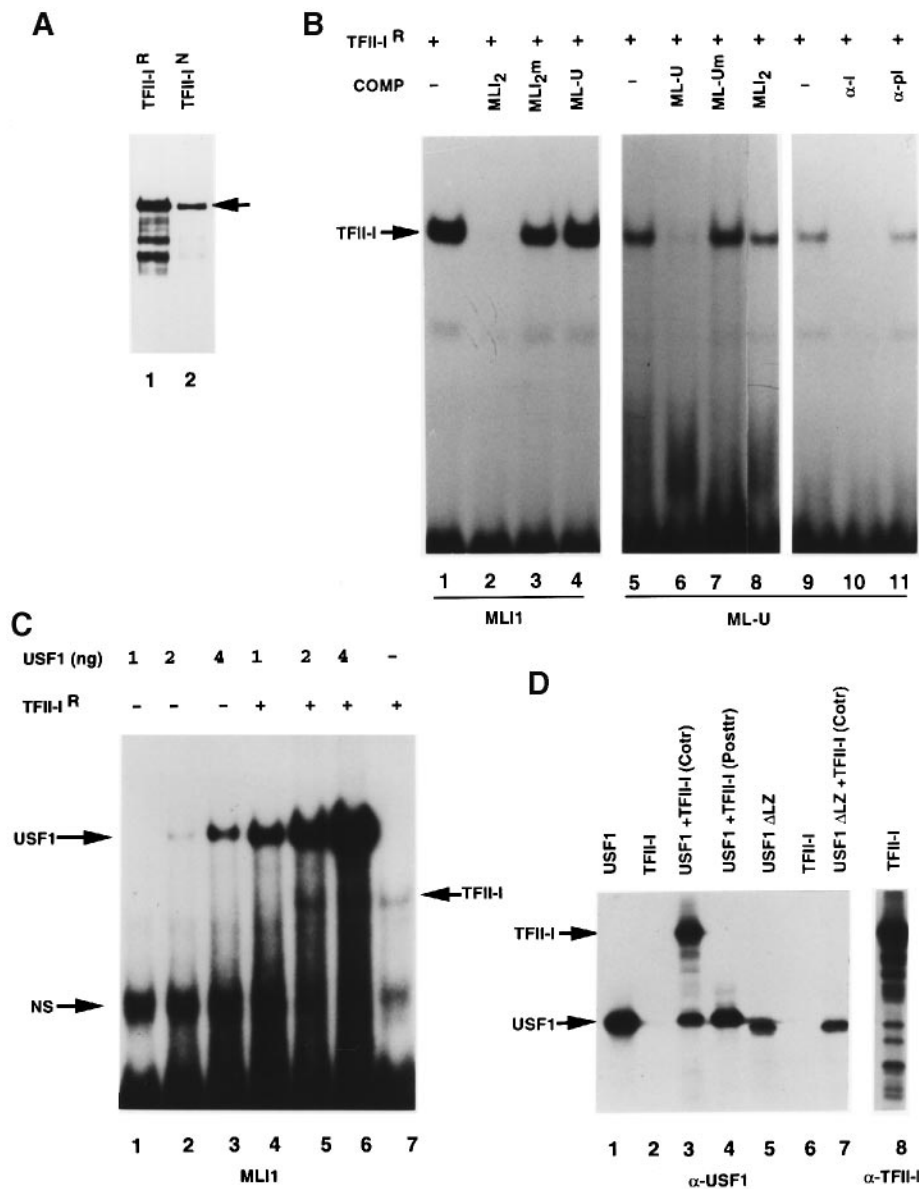
Finally, we tested the expression pattern of TFII-I in various tissue types. Northern blot analysis in HeLa- and Namalwa-derived poly(A)<sup>+</sup> RNA revealed that TFII-I is expressed as a single 4.7 kb message under stringent hybridization conditions (Figure 2D, left panel). Furthermore, as expected, a multiple tissue Northern blot analysis also showed that TFII-I is widely expressed (consistent with Western blot analyses, data not shown), although the extent of expression varied among different tissues (Figure 2D, right panel). Curiously, in these primary tissue types, in addition to the 4.7 kb TFII-I RNA, a shorter RNA at 4.2 kb was also visible. The structure of this RNA is unclear at present.

#### **Expression of a recombinant TFII-I that is competent in DNA binding**

For further functional tests, the cDNA encoding TFII-I was expressed via a bacterial expression vector that adds a hexa-histidine tag to the N-terminus of the protein, and recombinant protein was purified from crude bacterial lysate on a Ni<sup>2+</sup>-agarose column. A Western blot analysis with antibody raised against the putative DNA-binding domain (basic region, see above) of TFII-I (Figure 3A) showed a dominant 120 kDa band and several degradation products in the bacterially expressed recombinant TFII-I (TFII-I<sup>R</sup>, lane 1) in comparison with a single 120 kDa band (arrow) in native purified TFII-I (TFII-I<sup>N</sup>, lane 2). The anti-TFII-I antibody also recognized 120 kDa/TFII-I in various nuclear extracts (data not shown, and Manzano-Winkler *et al.*, 1996). Most importantly, as revealed by an EMSA with an oligonucleotide probe (MLI1) con-

taining Inr1, the recombinant TFII-I showed site-specific binding to AdML initiator elements (Figure 3B); the observed complex (lane 1) was shown to be specific by virtue of competition with intact MLI1- and MLI2-containing oligonucleotides (lane 2 and data not shown), but not with a mutant MLI2 oligonucleotide (lane 3). Furthermore, the binding of recombinant TFII-I to the Inr site was not competed by an E-box-containing oligonucleotide (lane 4). Finally, an EMSA with an oligonucleotide probe (ML-U) containing the AdML E-box demonstrated specific and direct binding of recombinant TFII-I to this element; the observed complex (lane 5) was competed by an oligonucleotide containing a wild-type E-box (lane 6), but not by an oligonucleotide containing a mutant E-box (lane 7) and only weakly by an Inr-containing oligonucleotide (lane 8). However, this binding could be inhibited specifically by an anti-TFII-I antibody (lanes 9–11, Figure 3B). Therefore, the Inr- and E-box-binding properties described here for the recombinant 120 kDa protein mirror those described for the native TFII-I (Roy *et al.*, 1991).

Having established intrinsic DNA-binding properties of recombinant TFII-I, we next tested its ability to interact with USF1 both on DNA (Figure 3C) and in solution (Figure 3D). As shown in Figure 3C, and consistent with previous studies of native TFII-I (Roy *et al.*, 1991), recombinant TFII-I significantly stimulated the binding of recombinant USF1 to the AdML-derived Inr element; recombinant TFII-I, like native TFII-I (Roy *et al.*, 1991), also stimulated USF1 binding to the E-box (data not shown). However, contrary to expectations of heterodimer formation, we were unable to observe a stable heterodimeric complex consisting of both TFII-I and USF1 under these conditions. This may reflect either an instability of the heterodimeric complex under the electrophoretic conditions employed or a role for TFII-I in increasing the stability of USF1 on Inr and E-box elements via transient interactions. In order to test whether TFII-I and USF1 can interact stably under different conditions, we performed co-immunoprecipitation studies subsequent to ectopic expression and radiolabeling of both proteins in a rabbit reticulocyte lysate. Under these conditions, roughly equivalent amounts of USF1 and TFII-I were synthesized, as shown by direct analysis of radiolabeled proteins (data not shown). In the immunoprecipitation analysis (Figure 3D), USF1 (lane 1) but not TFII-I (lane 2) was immunoprecipitated by an anti-USF antibody when these proteins were expressed separately. In contrast, TFII-I and USF1 were co-immunoprecipitated by anti-USF1 antibody when both proteins were co-translated (lane 3). TFII-I was not co-immunoprecipitated by anti-USF1 antibody when the two proteins were post-translationally mixed (lane 4). As a further control, co-immunoprecipitation was performed following expression of a mutant USF1 protein that lacked the LZ domain (lane 5). In this case, anti-USF1 antibody failed to co-immunoprecipitate TFII-I even when both proteins were co-translated (lane 7). That TFII-I is translated efficiently under these conditions was also demonstrated by immunoprecipitation of TFII-I by an anti-TFII-I antibody (lane 8). Taken together, these data demonstrate that stable interactions between USF1 and TFII-I do occur when the proteins are allowed to fold together. Thus, the inability to detect a stable heteromeric complex on DNA may reflect either the failure of the



**Fig. 3.** Analysis of recombinant TFII-I expressed in bacteria. (A) Western blot analysis of recombinant TFII-I (TFII-I<sup>R</sup>, lane 1) and the HPLC-purified native TFII-I (TFII-I<sup>N</sup>, lane 2). The arrow shows the 120 kDa polypeptide. (B) Specific binding of recombinant TFII-I to AdML Inr and E-box elements. Binding was monitored by EMSA with an AdML probe (MLI1) containing Inr1 (lanes 1–4) and with an AdML probe (ML-U) containing an E-box (lanes 5–11). Oligonucleotide competitors added at 50-fold molar excess contained: wild-type Inr2 sequences (MLI<sub>2</sub>), lanes 2 and 8; mutated Inr2 sequences (MLI<sub>2</sub><sup>m</sup>), lane 3; wild-type E-box sequences (ML-U), lanes 4 and 6; and mutated E-box sequences (ML-U<sup>m</sup>), lane 7. Anti-TFII-I serum (α-I) and pre-immune serum (α-pI) were added in lanes 10 and 11, respectively. (C) Stimulatory effect of recombinant TFII-I on USF1 binding to the AdML Inr1-containing probe (MLI1). The binding of variable amounts of recombinant USF1 was monitored by EMSA in the absence (lanes 1–3) and presence (lanes 4–6) of a fixed amount of recombinant TFII-I, which was also analyzed in the absence of USF1 (lane 7). (D) Interactions of *in vitro* translated USF1 and TFII-I in the absence of DNA binding. Intact TFII-I and both wild-type USF1 and a USF1 mutant lacking the leucine zipper (USF1<sup>ΔLZ</sup>) were co-translated in rabbit reticulocyte lysates in the presence of [<sup>35</sup>S]methionine, both individually and in the combinations indicated above the lanes. Individual translation reactions (lanes 1–3, 5–7 and 8), as well as a mixture of independently translated TFII-I and USF1 (lane 4), were subjected to immunoprecipitation with anti-USF1 antibody (lanes 1–7) or with anti-TFII-I antibody (lane 8). Immunoprecipitations were subjected to SDS-PAGE and autoradiography. Direct analyses of translation reactions revealed that approximately equal amounts of radiolabeled TFII-I and USF1 were synthesized when the corresponding vectors were expressed independently or together (data not shown).

independently synthesized proteins to interact stably, possibly because of improper folding, or the dissociation of TFII-I from the complex under the electrophoretic conditions. Although initial experiments have failed to detect formation of such a complex with co-translated USF1 and TFII-I (data not shown), this may reflect insufficient levels of synthesis in the *in vitro* system.

Although the recombinant TFII-I behaved similarly to

native TFII-I with respect to DNA-binding specificity and USF1 interactions, it did not show the *in vitro* transcription activity observed earlier (Roy *et al.*, 1991, 1993a) for native TFII-I preparations (data not shown). Thus, while confirming the DNA-binding specificity of the cDNA-encoded protein, these results also raise the possibility that the bacterially expressed TFII-I is improperly folded and/or lacking post-translational modifications that play a

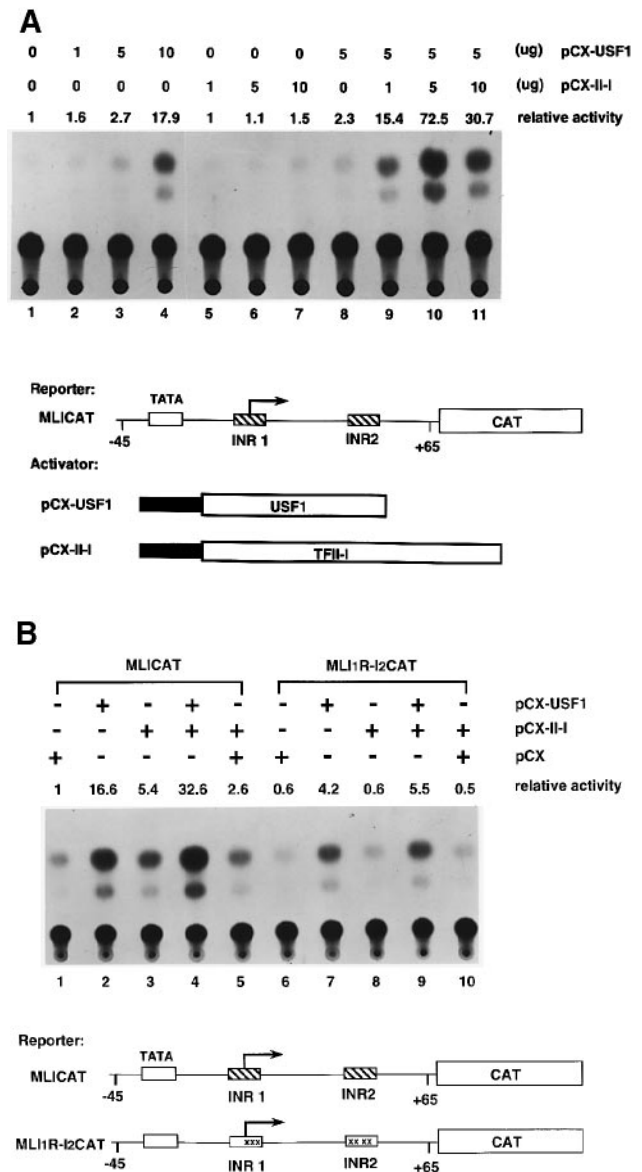
critical role in effecting the transcription function, but not the intrinsic DNA-binding activity. Another possibility is that although the 120 kDa polypeptide is competent in DNA binding, the transcriptional activity, as seen with the partially purified TFII-I, may reflect additional polypeptides associated with the 120 kDa polypeptide.

### Independent and synergistic functions of TFII-I and USF1 via the AdML *Inr* element in transfected cells

We reasoned that whether the inactivity of recombinant TFII-I in an *in vitro* transcription assay results from a lack of post-translational modifications or a lack of associated polypeptides, an analysis of TFII-I function in eukaryotic cell lines by transient transfection assays might circumvent these problems and reveal associated transcription functions. Furthermore, these assays would also enable us to test whether the synergism between TFII-I and USF1, as seen at the DNA-binding level, is manifested at the transcriptional level in a physiological situation.

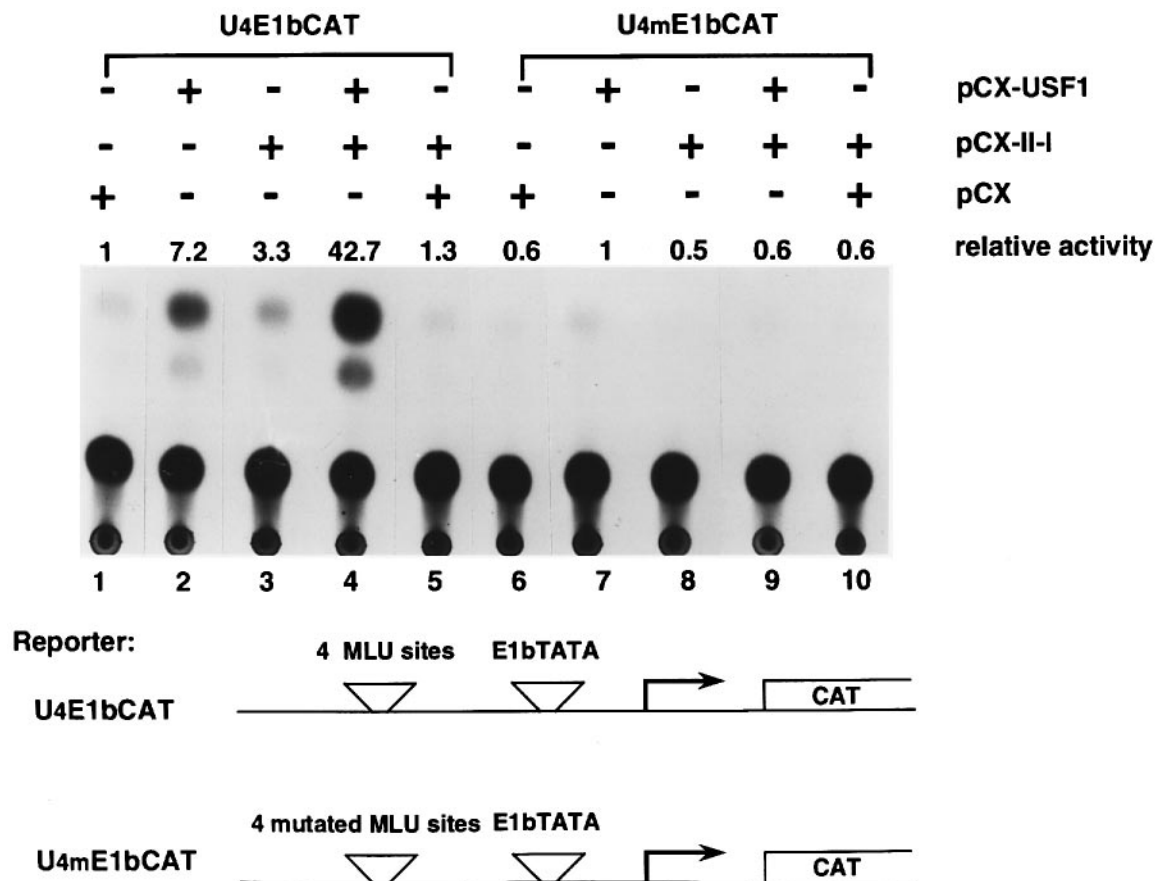
In order to study activation via AdML *Inr* sites, we used a reporter plasmid (MLICAT) containing the AdML core promoter (-45 to +65) fused to the CAT gene (Du *et al.*, 1993). In addition to the TATA element, the core promoter contains initiator elements at positions -3 to +9 (*Inr* 1) and +45 to +57 (*Inr* 2) (Roy *et al.*, 1991). HeLa cells were co-transfected with MLICAT and either an empty vector (pCX) or vectors expressing TFII-I (pCX-II-I) and/or the human USF1 (pCX-USF1) (Figure 4A). The reporter was activated significantly (up to 18-fold) by ectopic USF1 in a dose-dependent manner (lanes 1-4), but only marginally (1- to 1.5-fold) by ectopic TFII-I expression (lanes 5-7 versus lane 1). In contrast, at an intermediate level of USF1 expression (5  $\mu$ g of pCX-USF1) that gave only a 3-fold increase in reporter activity, co-expression of TFII-I resulted in markedly enhanced levels of activity that were up to 73-fold above the control values (lanes 9-11 versus lane 1). At the highest level of activity (lane 10), the overall activity was 25-fold greater than that expected (on the basis of additivity) from the independent expression of comparable levels of USF1 (lane 3) and TFII-I (lane 6). Greater than additive levels of activity were also observed at higher and lower levels of TFII-I co-expression with USF1 (lane 9 versus lanes 3 and 5, and lane 11 versus lanes 3 and 8). Hence, the effects of ectopic USF1 and TFII-I are clearly synergistic.

To test whether the transcriptional synergism between USF1 and TFII-I was *Inr* dependent, reporter plasmids containing either wild-type or mutant *Inr*1 and *Inr*2 core promoters were co-transfected with TFII-I or TFII-I plus USF1 (Figure 4B). Under the higher efficiency co-transfection conditions of this analysis (see legend to Figure 4B), the independent levels of activation by TFII-I and USF1 were slightly higher. Thus, the reporter was activated 17-fold by USF1 alone (lane 2), 5-fold by TFII-I alone (lane 3) and 33-fold by comparable concentrations of both together (lane 4). Although the greater effects of independently expressed USF1 and TFII-I resulted in a level of synergism lower than that observed in the analysis of Figure 4A, this did permit an analysis of the effect of *Inr* mutations on both independent and synergistic effects of ectopic USF1 and TFII-I. Consistent with previous reports (Du *et al.*, 1993), *Inr*1 and *Inr*2 mutations in the reporter template (MLI1R-I2CAT) reduced USF1-



**Fig. 4.** Independent and synergistic effects of ectopic USF1 and TFII-I expression on transcriptional activation through *Inr* elements in transfected cells. **(A)** Activation from the AdML core promoter. HeLa cells were co-transfected with the wild-type MLICAT reporter and with variable amounts of pCX-USF1 and pCX-II-I expression vectors, both alone and in combination, as indicated at the top of the figure. Transfection conditions were as described in Du *et al.* (1993). **(B)** Activation from AdML core promoters containing intact versus mutated *Inr*1 and *Inr*2 elements. HeLa cells were co-transfected with wild-type or *Inr*-mutated MLICAT reporters and with the indicated combinations of the pCX-USF1 expression vector (5  $\mu$ g), the pCX-II-I expression vector (5  $\mu$ g) or the control pCX vector (5 or 10  $\mu$ g, to bring the amount of total transfected DNA to 10  $\mu$ g). Transfection conditions were as described by Chen and Okayama (1987) and resulted in slightly higher transfection efficiency than the method used in Figure (A) and in Figures 5-7. In both (A) and (B), the relative CAT activities were normalized to the level of activity observed with the control pCX vector alone (lane 1) and are indicated above the figure.

mediated activation significantly (~3-fold) but not completely (Figure 4B, lanes 6 and 7 versus 1 and 2), possibly because of only partial inactivation of the *Inr* sites or of cryptic USF1-binding sites that function through the strong TATA element. More significantly, the *Inr*1 and *Inr*2



**Fig. 5.** Independent and synergistic effects of ectopic USF1 and TFII-I expression on transcriptional activation through E-box elements in transfected cells. HeLa cells were co-transfected with E1bCAT reporters containing intact versus mutated E-boxes and with the indicated combinations of the pCX-USF1 expression vector (5  $\mu$ g), the pCX-II-I expression vector (5  $\mu$ g) and the pCX control vector (5 or 10  $\mu$ g, to bring the total amount of transfected DNA to 10  $\mu$ g). The levels of CAT activities relative to that observed with the control pCX vector (lane 1), which reflects the function of endogenous factors, are indicated at the top of the figure.

mutations virtually eliminated both the TFII-I-mediated activation (lanes 8 versus 6, and 3 versus 1) and the synergism between USF1 and TFII-I (lanes 9 versus 7, and 4 versus 2); thus, the level of activity with TFII-I was equivalent to that observed with the control vector, and the level of activity with USF1 plus TFII-I was equivalent to that observed with USF1 alone. Although the sites of initiation in response to TFII-I have not been examined, earlier studies showed that the site of initiation was unchanged in response to USF1 and that the Inr mutations used here do not alter the site of initiation or the response of the mutated core promoter to activation by the SV40 enhancer (Du *et al.*, 1993). Thus, while the present results do not rule out functional interactions of USF1 or TFII-I with other sites in the promoter, they clearly show an independent effect of ectopic TFII-I and synergistic effects of ectopic TFII-I and USF1 that are dependent upon intact Inr elements.

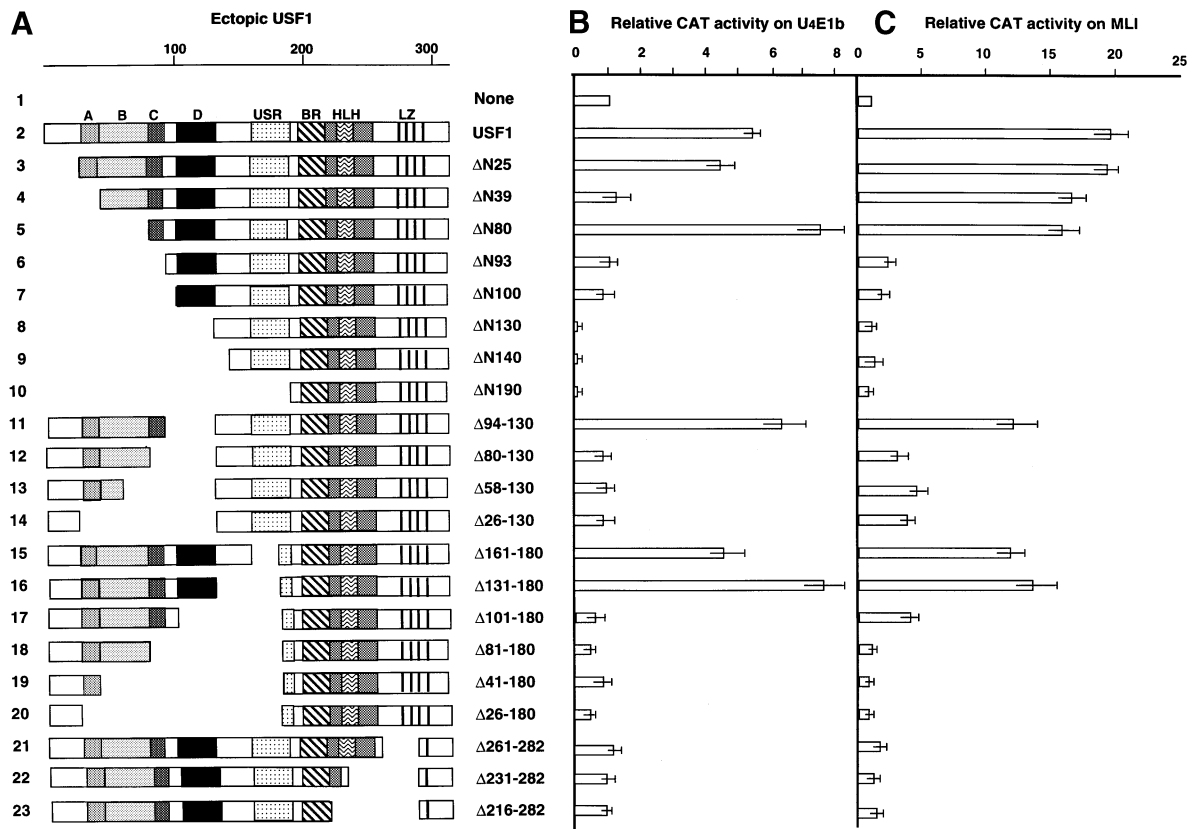
#### **Independent and synergistic interactions of TFII-I and USF1 via the AdML E-box in transfected cells**

TFII-I can bind independently not only to Inr elements but also to the upstream E-box (ML-U site) on the AdML promoter (Roy *et al.*, 1991). Therefore, potential activation functions of TFII-I through the ML-U site, in the absence of Inr elements, were analyzed *in vivo* using the U4E1bCAT reporter plasmid (Figure 5). This reporter contains four

ML-U sites (E-boxes), the adenovirus E1b TATA box region (but not the natural E1b Inr) and the CAT gene (Du *et al.*, 1993). Under the transfection conditions utilized (Figure 5), the reporter was activated 3-fold by TFII-I alone (lane 3 versus lane 1), 7-fold by USF1 alone (lane 2 versus lane 1) and 43-fold by USF1 and TFII-I together (lane 4 versus lane 1). The site specificity of transcriptional activation was demonstrated by employing a template containing mutations in the ML-U sites that block USF1 and TFII-I binding (Roy *et al.*, 1991; Du *et al.*, 1993). With this mutated reporter plasmid, neither USF1 nor TFII-I exhibited any independent or synergistic activation function (Figure 5, lanes 6–10). These results confirm the previous demonstration that USF1 can activate transcription *in vivo* through E-box elements (Du *et al.*, 1993) and further show (i) that ectopic TFII-I also can activate transcription *in vivo* through E-box elements in the apparent absence of an Inr site, and (ii) that ectopic USF1 and TFII-I can function synergistically through E-box sites to activate transcription *in vivo*.

#### **USF1 domains responsible for E-box-dependent activation**

To delineate the USF1 domain(s) responsible for its activation functions via the E-box (ML-U) and Inr (MLI) sites, as well as its synergistic functions with TFII-I *in vivo*, 21 USF1 deletion constructs were generated



**Fig. 6.** USF1 domains important for activation through E-box and Inr elements in transfected cells. **(A)** Schematic diagram of USF1 and derived mutants. Present or previously described domains in intact USF1 (line 2) include the basic region (BR), the helix-loop-helix region (HLH), the leucine zipper region (LZ), the conditionally required USR domain, the potent activation domain A, the negative regulatory domain B and its counteracting activation domain A, and the spacer/activation domain D (for discussion and references, see text). The bar at the top indicates positions within the 310 residue USF1. The numbers in the names of each mutant, indicated schematically in lines 2–23, indicate the numbers of the amino acids deleted from the N-terminus or from internal regions. **(B)** Transactivation of the E-box-containing U4E1bCAT reporter. **(C)** Transactivation of the Inr-containing MLI-CAT reporter. In **(B)** and **(C)**, cells were co-transfected with either the U4E1bCAT **(B)** or the MLI-CAT **(C)** reporter and with 5  $\mu$ g of either the control pCX vector (line 1) or the indicated pCX-USF1 expression vectors (lines 2–23). The horizontal bars show the levels of activation by ectopic USF1 expression relative to the level of activity due to endogenous factors (line 1).

(Figure 6A) and analyzed in transfection assays with the U4E1bCAT reporter (Figure 6B). The levels and stabilities of these ectopically expressed truncated proteins in HeLa cells were monitored and normalized by Western blot analysis and were not found to vary significantly (data not shown).

As reported previously, intact USF1 specifically binds to DNA as a dimer and requires the LZ and the HLH regions for such binding (Gregor *et al.*, 1990). As expected, mutants lacking part of the LZ ( $\Delta$ 261–282), the LZ plus the second helix of the HLH region ( $\Delta$ 231–282) or the LZ plus the entire HLH region ( $\Delta$ 216–282) showed no activation function relative to intact USF1 (Figure 6B, lines 21–23 versus line 2), which activated the reporter 5-fold above the control level (line 2 versus line 1). Hence dimerization and stable DNA binding are essential for transcriptional activity of USF1 via the E-box *in vivo*, as well as *in vitro* (Kirschbaum *et al.*, 1992).

Analysis of mutations outside the DNA-binding and dimerization domains revealed regions important for activation *per se*. Deletion of 25 amino acids from the N-terminus ( $\Delta$ N25) had little effect on USF1 activation through the ML-U site, whereas further deletion to residue 39 ( $\Delta$ N39) resulted in a marked decrease in activity to the control level (lines 3 and 4 versus lines 1 and 2).

Surprisingly, deletion to position 80 resulted in a slightly increased activity compared with the wild-type level (line 5), whereas a further deletion to position 93 or 100 resulted in a dramatic decrease in the transcriptional activation (lines 6 and 7). Deletions that extend to positions 130, 140 and 190, but maintain the bHLH-LZ region, resulted in a repression of endogenous USF activity on the reporter plasmid (lines 8–10). Taken together, these studies indicate the presence of at least two activation domains, one located between residues 25 and 40 (domain A) and another between residues 80 and 93 (domain C). While domain A appears to be conditionally required to counteract the adjacent inhibitory domain B, domain C appears to be a very potent and more conventional activation domain that functions in the absence of both domains A and B. In addition, the region from residues 100 to 130 (domain D) contributes to the activity in such a way that removal of this region results in the transition from an inactive to a dominant-negative mutant. Hence, this domain may reduce potentially stronger interactions of  $\Delta$ N130 USF1 with E-boxes or with intact USF1 that directly (through heterodimerization) or indirectly (through E-box binding) inhibit the function of endogenous USF, but whether it can be regarded as a conventional activation domain is not clear (see also below). Finally, the region



adjacent to the C-terminus of domain D appears to impart some inhibitory effects since an internal deletion removing this region resulted in a moderate (40%) increase in transcriptional activity compared with the wild-type USF1 (line 16). The other internal deletions had effects largely consistent with these conclusions, with those deletions removing domain C resulting in a near complete loss of activity (lines 12–14 and 18–20). The very significant loss of activity following deletion of domain D from the active  $\Delta$ 131–180 mutant (lines 17 versus 16) suggests the possibility of an activation domain that could function in cooperation with domain C, but the retention of full USF1 activity with a simple deletion of this region (line 11) suggests that it may simply act as an essential spacer between domain C and the DNA-binding domain. Also of note is the failure to observe an activation domain (USR) between residues 158 and 188 that has been reported by Luo and Sawadogo (1996) to be evolutionarily conserved between USF1 and USF2 and active in the absence of N-terminal domains on the AdML promoter. This difference most probably reflects the interesting observation that USR function requires an Inr element, not present in the U4E1bCAT, acting in synergy with the E-box (Luo and Sawadogo, 1996).

#### **USF1 domains responsible for Inr-dependent activation**

To map the USF1 domains responsible for activation through the Inr site, we used the above described USF1 mutants in conjunction with the reporter plasmid (MLICAT) containing the AdML core promoter (–45 to +65) fused to the CAT gene (Du *et al.*, 1993). Consistent with earlier described results, ectopic expression of USF1 activated the reporter ~18-fold above the level resulting from endogenous factors (Figure 6C, line 2 versus line 1). As expected, and consistent with the importance of the LZ and HLH regions in mediating specific DNA binding and activation via the E-box (above), deletions of these regions resulted in a complete loss of transactivation function via the Inr element as well (lines 21–23 versus line 2). N-terminal deletions removing domains A and B had little effect on USF1 function (lines 3 and 4 versus line 2), indicating that these domains are not required for, and do not influence, the independent function of ectopic USF1 through Inr elements; this contrasts with the results (above) observed for USF1 function through the E-box element. However, deletions removing domain C had drastic effects on USF1 function through the Inr (line 6 versus line 5, and line 12 versus line 11). Hence, domain C appears critical for activation by USF1 not only through E-boxes (above) but also through Inr elements. The results observed with other (internal) deletion mutants (lines 11–15 and 18–20 versus line 2) are in accord with the major importance of domain C for USF1 function through Inr elements, again with an apparent spacing requirement (line 17 versus lines 16 and 11). As described for USF1 function via E-boxes on the E1b promoter, there was also little activity on the MLI reporter of USF1 mutants containing an intact USR region but lacking domains A–D (lines 8–9). This may reflect a requirement, for optimal USR function, for both E-box and Inr elements (Luo and Sawadogo, 1996). The modest (40%) reduction of activity with an internal deletion removing part of the USR (line

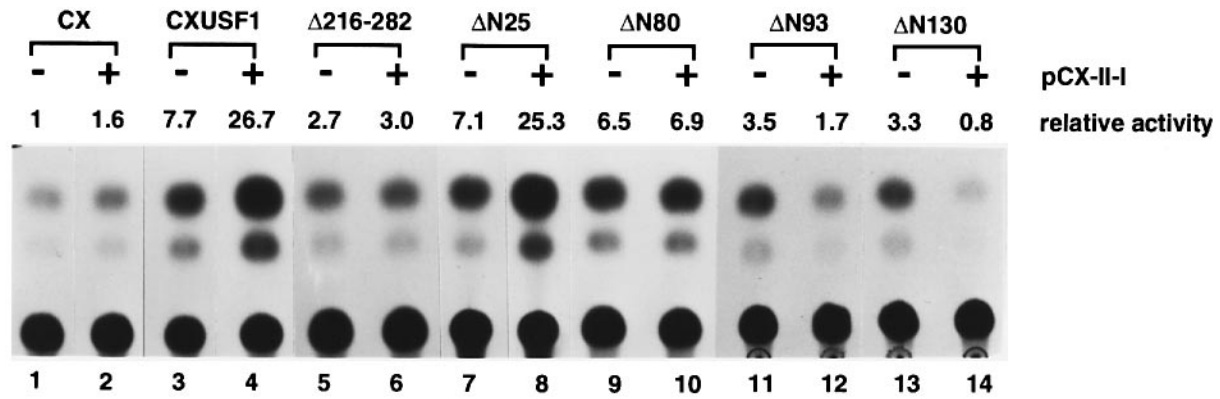
15 versus line 2) may suggest a partial function of the USR through Inr elements.

#### **Synergistic activation of USF1 with TFII-I via the Inr elements requires both protein–protein interaction and activation domains of USF1**

In order to map the interaction domain(s) necessary to mediate transcriptional synergy with TFII-I via the AdML Inr elements, five USF1 mutants were analyzed in transfection assays with the MLICAT reporter (Figure 7). The  $\Delta$ 215–283 mutant, which lacks the LZ and the HLH regions but contains the basic region, showed no detectable transcriptional synergy with TFII-I (lanes 5 and 6). The  $\Delta$ N25 mutant, which lacks the first 25 amino acids of the N-terminus and shows the wild-type level of activation via the MLI site (Figure 6C), exhibited a level of transcription synergy with TFII-I similar to that observed with wild-type USF1 (lanes 7 and 8 versus lanes 2 and 3). However, the  $\Delta$ N80 mutant, which lacks the first 79 amino acids of the N-terminus (including domains A and B) and shows a wild-type level of transactivation by itself on the MLICAT reporter (Figure 6C), exhibited no transcriptional synergy with TFII-I. In addition, the  $\Delta$ N93 and  $\Delta$ N130 mutants, which lack activation domains C and C plus D, respectively, but contain the bHLH-LZ domain, behaved as dominant negatives in the presence of TFII-I. Together, these results suggest (i) that the USF1 dimerization (HLH-LZ) domain and at least one N-terminal activation domain are both required for synergistic effects with TFII-I and (ii) based on the TFII-I-dependent dominant-negative effects, the possibility that USF1 and TFII-I may form stable heteromeric complexes *in vivo* at the Inr. Consistent with the latter suggestion, and indicative of TFII-I interactions with USF1 through the DNA-binding/dimerization domain, TFII-I stimulated binding to the Inr of USF1 mutants lacking the region upstream of the DNA-binding/dimerization domain (data not shown).

## **Discussion**

Although transcriptional regulation involves communication between activators bound to distal control elements and general factors acting through core promoter elements, the mechanisms involved in regulating broad classes of natural promoters are not well understood. Thus, while there is considerable information on activator interactions with general initiation factors and associated cofactors (reviewed in Orphanides *et al.*, 1996; Roeder, 1996; Verrijzer and Tjian, 1996; Ptashne and Gann, 1997) that function through TATA elements, there is little information on comparable mechanisms of activation through Inr elements that function alone or in concert with TATA elements and that require novel basal factors (Roeder, 1996). Adding to this complexity are observations that certain activators may function selectively on specific core promoter configurations (reviewed in Novina and Roy, 1996; Smale, 1997). Toward a further analysis of this problem, the present study has focused on two factors, TFII-I and USF1, that interact physically and functionally through both upstream regulatory (E-box) and core promoter (Inr) elements. These studies have been facilitated by the purification, cognate cDNA cloning and ectopic expression of TFII-I, and clearly demonstrate a role for



**Fig. 7.** The synergistic action of TFII-I and USF on transcriptional activation *in vivo* requires both activation and dimerization domains of USF1. HeLa cells were co-transfected with the MLCAT reporter (5  $\mu$ g), an expression vector for wild-type or mutant forms of USF1 (5  $\mu$ g) and either an expression vector for TFII-I (5  $\mu$ g) or the control pCX vector (5  $\mu$ g), as indicated above the figure. CAT activities relative to the activity observed with endogenous factors in the absence of ectopic USF1 or TFII-I (lane 1) are indicated at the top.

TFII-I, as a novel co-regulator for USF1 *in vivo*. They also provide further evidence for apparently dual functions for both TFII-I and USF1 at both distal E-boxes and at Inr elements, as well as insights into potentially integrated functions during communication between these elements.

#### TFII-I structure

The purification of native TFII-I on the basis of its Inr-binding properties permitted the cloning of a cDNA encoding a 120 kDa polypeptide whose relationship to native TFII-I was established by DNA binding and immunological assays (below). Perhaps the most interesting aspect of the TFII-I sequence is the presence of six highly conserved 90 residue repeats. A striking feature of these repeats is the presence of potential HLH-like domains that have been implicated in homo- and hetero-dimerization of conventional HLH proteins (reviewed in Ferre-D'Amare *et al.*, 1993). Although the 'loop' domain is larger than usually observed, flexibility in the length of this domain can be accommodated in the 3-D co-crystal structure of HLH proteins (Ferre-D'Amare *et al.*, 1993), thus allowing, in principle, a similar structure for the HLH-like domains in TFII-I (S.K.Burley, personal communication). Whether TFII-I behaves as a conventional bHLH protein with respect to specific protein-protein and protein-DNA interactions, and whether the individual repeats may show different specificities, remains to be determined. Nevertheless, the structure of TFII-I could potentially allow a variety of different protein-protein and protein-DNA interactions that explain many of the diverse properties of TFII-I (see further discussion below). Another interesting feature of the TFII-I sequence is the presence of a number of potential phosphorylation sites, including a consensus (PXSP) site for mitogen-activated protein kinase (MAPK) and several Src autophosphorylation sites (EDXDY), suggesting that the various phosphorylation modifications could increase the functional diversity of TFII-I. In agreement, recent studies have shown that a protein (designated BAP-135) identical to the 120 kDa TFII-I polypeptide both interacts with a tyrosine kinase (Btk) important for B cell activation and is phosphorylated on tyrosine in response to B cell receptor engagement *in vivo*, although a function for this protein was not established (Yang and Desiderio, 1997; C.D.Novina,

U.Bajpei, H.H.Wortis and A.L.Roy, unpublished observations).

#### *In vitro* interactions and functions of recombinant TFII-I

The cDNA cloning of the 120 kDa TFII-I polypeptide has also allowed us to test the various functions that were attributed originally to the native factor. The relationship of the cDNA-encoded polypeptide was verified by the ability of the encoded protein, expressed in bacteria, to show specific binding to Inr-containing probes, and by the ability of corresponding antibodies to inhibit specific binding of native TFII-I to Inr elements (Manzano-Winkler *et al.*, 1996). Like the native protein, the bacterially expressed recombinant protein also demonstrated specific binding not only to the Inr elements, but also to the upstream E-box of the AdML promoter. Furthermore, the binding of TFII-I to one site could not be competed with oligonucleotides containing the other binding site, suggesting two distinct DNA-binding domains and thus the potential of simultaneous binding to both sites. The recombinant TFII-I also markedly enhanced the binding of USF1 to the AdML Inr element (which is otherwise not a high affinity USF1-binding site), although a stable heteromeric complex between the two proteins on DNA was not observed. The possibility that this reflects an instability of the complex under electrophoretic conditions, rather than a catalytic effect of TFII-I, is suggested by co-immunoprecipitation experiments that indicate a stable interaction between the two proteins under appropriate folding conditions (e.g. co-translation). Regardless of whether the effects of TFII-I on USF1 binding are transient or stable, these results suggest that the intrinsically weak interactions of USF1 are greatly accentuated in the presence of TFII-I. They further suggest a novel gene regulatory mechanism involving USF1 and TFII-I functions that are dependent upon the Inr element. At the same time, a significant but less dramatic effect of recombinant TFII-I on the binding of USF1 to the AdML E-box was also observed (data not shown, Roy *et al.*, 1991), consistent with a potentially related function at E-boxes.

Although the bacterially expressed TFII-I showed the same DNA-binding specificity and interactions with USF1 as did native TFII-I, the recombinant protein failed to

show the Inr-dependent transcription activity manifested by native TFII-I in a partially purified reconstituted system (Roy *et al.*, 1991, 1993a). This could reflect improper folding, the lack of natural post-translational modifications or the lack of associated polypeptides present in native TFII-I. At the same time, we note that the previously demonstrated function of native TFII-I was observed in a system reconstituted with TBP and other partially purified general factors; in the absence of TFII-I, this system was dependent upon TFIIA for activity (Roy *et al.*, 1993a). These observations are surprising in retrospect, since TATA-directed basal transcription in highly purified systems reconstituted with TBP does not generally require TFIIA (or TFII-I) (Orphanides *et al.*, 1996; Roeder, 1996), but may have reflected the limiting levels of TBP employed in the assay and the presence of TBP-interacting negative cofactors whose effects can be reversed by TFIIA (Meisterernst and Roeder, 1991) and potentially by TFII-I. Further, while TBP is sufficient to promote basal transcription from a TATA-containing promoter (Roeder, 1996), both TBP and TAFs are required for Inr-driven basal transcription—both in the absence (Martinez *et al.*, 1994, 1995) and presence (Martinez *et al.*, 1994; Verrijzer *et al.*, 1995; Kaufman *et al.*, 1996) of a TATA element. In view of these complications, and as a prelude to *in vitro* studies with recombinant TFII-I expressed in eukaryotic cells, we have relied on transfection studies to document transcriptional functions for the cloned TFII-I component.

#### ***In vivo* function and structural features of USF1**

Toward a further understanding of the function of USF1 through Inr and E-box elements, and ultimately the synergy with TFII-I, functional domains were mapped in transfection assays. Apart from the previously described bHLH-LZ DNA-binding and dimerization domain, four other domains were implicated in these *in vivo* functions of USF1 (Figure 6). Domain A acts to counteract a negative effect of the adjacent modulatory domain B, while domain C is a very potent and more conventional activation domain that could account for much of the activation by USF1. Domain D, while potentially representing an activation domain, may function mainly as an essential spacer between the potent C domain and the DNA-binding domain. Two of these domains, notably A and D, were also revealed in previous *in vitro* assays with an E-box-containing reporter different from that used here (Kirschbaum *et al.*, 1992).

Interestingly, the activation domains essential for USF1-mediated activation through E-box and Inr elements are not identical. Thus, whereas the bHLH-LZ DNA-binding domain and the potent activation domain C were critical for activation from both elements, effects of the A and B domains were apparent only through the E-box elements when USF1 was analyzed in the absence of ectopic TFII-I. In addition, USF1 mutants lacking the N-terminal region containing domains A–D showed dominant-negative effects (on endogenous USF) only through E-box and not through Inr elements. Thus, there appear to be both common features, as well as some differences, in the mechanisms of USF1 activation through distal E-box and proximal Inr elements. The differences, especially the lack of dominant-negative effects at the Inr, may reflect variations in USF1 binding at E-box versus Inr elements,

the positions of the E-box and Inr elements relative to the TATA element and its associated general factors, or the function of an additional cofactor at the Inr sites.

#### ***In vivo* function of TFII-I in cooperation with USF1**

Although moderate but significant effects of ectopic TFII-I expression on Inr-dependent transcription could be observed under high efficiency transfection conditions (Figure 4B), the effects under standard conditions (Figure 4A) were low—most probably reflecting a high endogenous level of TFII-I. In contrast, ectopic TFII-I expression consistently and dramatically enhanced ectopic USF1 function through both E-box and Inr elements. These results provide unequivocal evidence for an *in vivo* transcription function, through Inr and E-box elements, for TFII-I.

Although the mechanistic basis for the functional synergy between TFII-I and USF1 is not yet well understood, it may be explained in part by the observed physical interactions in solution and the synergism in DNA binding. The synergistic DNA-binding effects are dependent only upon the DNA-binding and dimerization domains of USF1, whereas synergistic transcription functions require USF1 activation domains. Indeed, a USF1 mutant lacking any activation domain acts as a dominant-negative in the presence of TFII-I, possibly reflecting enhanced USF1–TFII-I interactions at promoter sites and consistent with stronger effects of TFII-I on binding of activation domain-deficient USF1 mutants than on intact USF1 in EMSA (data not shown). Interestingly, the functional synergism requires USF1 domains (A and B) that are not essential for the independent function of ectopic USF1, although it is possible that the potent activation domain C (required in all other functional tests) is necessary as well. These results are reminiscent of others showing physical interactions between tissue-specific bHLH proteins and essential co-regulators through DNA-binding/dimerization domains and functional synergy dependent on associated activation domains (Molkentin and Olsen, 1996). Thus, we favor a model in which synergistic DNA-binding is mediated through interactions of the DNA-binding/dimerization domains, with subsequent interactions of activation domains in USF (and potentially TFII-I) interacting with components of the general transcription machinery. Although the actual targets in this case are unknown, previous studies have documented USF interactions with a partially purified TFIID (Sawadogo and Roeder, 1985) and with a specific TAF (Chiang and Roeder, 1995). The possibility of targets within factors (especially TFIID) assembled at TATA elements may explain the function of USF1 and TFII-I both at distal E-boxes and at proximal Inr sites in corresponding TATA-containing promoters.

#### ***Diverse and potentially interrelated functions of TFII-I***

While the above discussion has stressed potentially diverse USF1/TFII-I functions at E-box versus Inr elements, a number of TATA-containing and TATA-less promoters contain both E-box and Inr elements. In the case of the AdML promoter, genetic analyses with intact adenovirus have provided evidence for functional interactions of E-boxes with both Inr and TATA elements during virus

replication (Lu *et al.*, 1997), and similar E-box–Inr interactions are probable in the case of Inr-containing TATA-less promoters containing functional E-boxes (Outram and Owen, 1994). In conjunction with the present results, these studies raise the interesting possibility of synergistic mechanisms involving concomitant TFII-I–USF1 interactions at both E-box and Inr sites. Previous indications (Ferre-D'Amare *et al.*, 1994) that USF1 multimers can occupy two DNA sites simultaneously are in accord with this possibility, with the additional complexity of simultaneous TFII-I interactions serving either to help establish stable USF1–DNA complexes or to participate as a direct DNA-binding partner.

The present studies implicate TFII-I as a novel co-regulator for a specific bHLH factor (USF1), and such a co-regulator may help explain the differential functions of various ubiquitous and cell-specific bHLH proteins either in the absence or in the presence of cognate E-boxes (Molkentin *et al.*, 1995). At the same time, the possibility of TFII-I serving as a co-regulator for other types of activators is apparent from recent studies (Gruneberg *et al.*, 1997) showing that TFII-I can interact with and facilitate interactions of SRF and Phox proteins at distal elements in the *c-fos* gene. Interestingly, SRF belongs to the MADS family of transcription factors (Shore and Sharrocks, 1995), a member of which physically and functionally interacts with cell-specific bHLH proteins (Molkentin *et al.*, 1995). Therefore, despite being a ubiquitous factor, TFII-I might integrate diverse cell type-specific transcriptional responses by virtue of its interactions with both bHLH and MADS family proteins. The cDNA cloning and initial characterization of recombinant TFII-I sets the stage for a more detailed analysis of its functions, including its potential role as a determinant for Inr-mediated transcription initiation in specific promoters.

## Materials and methods

### Purification of TFII-I

TFII-I was purified from standard HeLa nuclear extract (Dignam *et al.*, 1983). The HeLa nuclear extract was dialyzed against buffer B [20 mM Tris (pH 7.9 at 4°C), 0.2 mM EDTA, 5 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 10% (v/v) glycerol] containing 100 mM KCl (B100) and fractionated by chromatography on phosphocellulose (Whatman, P11) according to standard procedures (Dignam *et al.*, 1983). For the preparation of TFII-I reported in these studies, the P11 0.3 M KCl fractions, containing most of the p120 polypeptide was generously provided by M.Meisterernst. The 0.3 M KCl fraction (150 mg of protein) was dialyzed against buffer B containing 40 mM KCl (B40) and loaded onto a DEAE-52 (Whatman) column (BioRad Econo-Column, 5×10 cm<sup>2</sup>, 125 ml bed volume). The column was developed with a 700 ml linear gradient of 40–350 mM KCl in buffer B at a flow rate of 2 ml/min. TFII-I protein eluted between 100 and 120 mM KCl. These fractions were pooled and loaded onto a Mono-S FPLC (HR 5/5, Pharmacia) column. The column was developed with a 10 column volume (10 ml) linear gradient of 100–500 mM KCl at a flow rate of 0.5 ml/min. The TFII-I peak, as determined by Inr element-binding activity, eluted between 300 and 315 mM KCl. These fractions were pooled, dialyzed against B100 and applied to a 1 ml bed volume double-stranded DNA cellulose (dsDNA, Sigma) column (0.64 cm × 5 cm, Bio-spin from BioRad). The column was washed successively at 15 column volumes/h with two column volumes of B100, three column volumes of B200 and three column volumes of B500. The TFII-I protein peak (determined by silver stain) co-eluted with the peak DNA-binding activity in fraction three of the 500 mM KCl fractions. Most experiments performed in this report employed these fractions following dialysis against B100.

For preparation of homogeneous TFII-I/p120, the dsDNA cellulose

fractions containing TFII-I were combined (1 mg), dialyzed against buffer C [20 mM HEPES, pH 7.9 at 25°C, 0.5 mM EDTA, 10% (v/v) glycerol] and loaded onto a SP-5PW HPLC column (75×7.5 mm, BIO GEL) at room temperature (25°C). The column was washed with 20 ml of buffer C containing 100 mM KCl and eluted with a (20 ml) linear gradient of 100–500 mM KCl in buffer C. Fractions containing pure TFII-I eluted between 300 and 330 mM KCl. Fifty pmol of this homogeneous preparation was subjected to SDS–PAGE, electroblotted onto nitrocellulose and submitted for microsequencing. Three of the derived peptide sequences were used to design best-guess oligonucleotide probes to screen the Namalwa cDNA library.

### cDNA screening

Peptide sequences obtained from microsequencing of the 120 kDa polypeptide included P10(PENYDLATK), P13(PELVI[R]YLPPI[T]MA) and P17(VIRPFPGLVINQLVDQ). Two residues in P13 were ambiguous and are indicated by parenthesis. A 'best-guess' oligonucleotide probe was synthesized for each polypeptide, using codons predicted by Lathe (1985). The oligo(dT)-primed Namalwa cDNA library was screened sequentially with each best-guess probe. Three clones were obtained by hybridization to the P10 probe (5'-CCT GAG AAC TAT GAC CTG GCC ACC CTG AAG-3') at low stringency (37°C). Secondary screens were done at 42°C. One clone with a 3 kb insert, 22.1, was confirmed by sequencing, and was found to encode all three peptides obtained from microsequencing of native TFII-I. An additional clone, 9.2, was obtained by rescreening the library with a fragment from 22.1 at high stringency. The 1.3 kb insert from 9.2 was sequenced, and contained 780 bp found in 22.1 and an additional 540 bp of 5' sequence. A full-length TFII-I cDNA (clone 3.1) was obtained by rescreening the library with the *Pst*I 5'-end fragment of clone 9.2.

### Northern blot analysis

Total cellular RNA from HeLa (7 µg) or Namalwa (15 µg) cells was resolved by agarose gel electrophoresis, transferred to a nitrocellulose membrane and hybridized with a 3'-end fragment (*Eco*RI–*Hind*III) of clone 22.1 that was labeled by random priming. For the multiple tissue Northern blot, a poly(A)<sup>+</sup> RNA-containing membrane (Clontech) was hybridized with a 1.5 kb 3'-end *Eco*RI fragment that was labeled by random priming.

### Bacterial expression of TFII-I

To express TFII-I in bacteria, the full-length TFII-I cDNA was subcloned into His<sub>6</sub>pET11-d (Invitrogen) at the *Nde*I–*Bam*HI site. The resulting plasmid, pET11-d-II-I, allowed overexpression of His<sub>6</sub>-TFII-I in the bacterial T7 expression system BL21(DE3)pLYS-S (Stratagene). BL21(DE3)pLYS-S cells containing the expression plasmids were grown overnight in 3 ml of LB media containing chloramphenicol and carbenicillin antibiotics. One ml of the overnight culture was added to 500 ml of TBM 9 media (10 g of Bacto tryptone, 4 g of glucose, 5 g of NaCl, 1 mM of MgSO<sub>4</sub>, 1 g of NH<sub>4</sub>Cl, 3 g of KH<sub>2</sub>PO<sub>4</sub>, 6 g of Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O per 1 l of media), grown to an A<sub>600</sub> of ~0.3 and induced with 0.1 mM of isopropyl-β-D-thiogalactopyranoside for 3 h at 30°C. After 3 h, the cells were harvested by centrifugation at 8000 r.p.m. for 15 min.

### Purification of bacterially expressed recombinant TFII-I

Conditioned medium was aspirated and the dry pellet was snap-frozen on dry ice. The pellet was resuspended in 30 ml of BC500 buffer [20 mM Tris–HCl, pH 7.9, 0.5 M NaCl, 20% glycerol, 1 mM PMSF, 1% aprotinin, leupeptin (1 µg/ml), soybean trypsin inhibitor (1 µg/ml), antipain (1 µg/ml), 20 mM β-mercaptoethanol and 0.1% NP-40] and subjected to sonication for 4 min on ice. The slurry was centrifuged and the supernatant was loaded onto a 1 ml ProBond Ni<sup>2+</sup>-agarose column (Invitrogen). The column was washed with 20 ml of BC500 lacking NP-40 and β-mercaptoethanol and eluted in BC500 supplemented with 200 mM imidazole. The pooled fractions were dialyzed against BC100 and used for DNA binding and interaction studies.

### Generation of anti-TFII-I polyclonal antibodies

Polyclonal anti-TFII-I antibodies were raised in rabbits (Research Genetics, Alabama) employing the synthetic peptide GKRKVREFNFKEK-WNARITDL, which corresponds to amino acid residues 301–321 in the hypothetical translation of the TFII-I cDNA.

### Western blot analysis

TFII-I proteins were subjected to 7.5% SDS–PAGE and transferred to nitrocellulose by the semi-dry blotting method in a buffer containing

0.192 M glycine, 0.025 M Tris base, 20% methanol. The blot was blocked in TBS (10 mM Tris pH8.0, 150 mM NaCl) with 6% non-fat dry milk (Carnation). For TFII-I Western blots, primary (anti-TFII-I, 1:2500 dilution) and secondary (1:1500 dilution) anti-rabbit horseradish peroxidase (HRP)-linked antibodies were incubated in TBS containing 0.05% Tween-20. All Western blots were developed by enhanced chemiluminescence (ECL, Amersham).

#### Electrophoretic mobility shift analysis (EMSA)

The EMSA reactions in Figure 3B were performed either with an AdML Inr1 probe (MLI1) containing sequences from -15 to +23 or with an AdML E-box probe (ML-U) containing sequences from -75 to -35 (Roy *et al.*, 1991). The EMSA reactions in Figure 3C were performed with a longer AdML Inr1 probe (MLI1) containing sequences from -22 to +43 (Roy *et al.*, 1991). The probes were labeled with [ $\alpha$ - $^{32}$ P]dCTP (3000 Ci/mmol) and Klenow fragment. The competitors were: MLI2 (AdML Inr2 oligonucleotide containing sequences from +24 to +67); MLI2m (MLI2 with mutations G→C at +42, G→C at +46, AG at +48, C→G at +53, T→G at +54, T→A at +57 and G→C at +58); ML-U (oligonucleotide corresponding to Ad2ML sequences from -75 to -35 and containing the E-box sequence CACGTG); and ML-Um (ML-U with the mutated E-box sequence CCCGAT). The competitors were also Klenow-filled with cold nucleotides. DNA-binding reactions proceeded for 20 min at 30°C. The final reaction volume was 20  $\mu$ l in buffer B containing 80 mM KCl, 5 mM DTT and 100 ng of poly(dA:dT) as non-specific competitor. All reactions were subjected to electrophoresis through a 5% native polyacrylamide gel containing 5% glycerol in 0.5 $\times$  TBE (40 mM Tris pH7.6, 40 mM boric acid, 2 mM EDTA) for 3 h at 140 V.

#### Co-immunoprecipitation

TFII-I and USF1 cDNAs were cloned into the pT7 expression vector (Gregor *et al.*, 1990). Templates (1  $\mu$ g) were linearized and used in an *in vitro* coupled transcription-translation reaction in TNT rabbit reticulocyte lysate (Promega) in the presence of 3  $\mu$ l of [ $^{35}$ S]methionine (Amersham). The TFII-I template contained the complete ORF and the USF1 template contained either the complete ORF or an ORF ( $\Delta$ LZ) missing the LZ but containing the HLH domain (Kirschbaum *et al.*, 1992). Two  $\mu$ l of each of these reactions were used for immunoprecipitation assays using either the anti-USF1 antibody raised against the full-length protein or an anti-TFII-I antibody. The assay was carried out in buffer B containing 0.1% NP-40. Precipitated proteins were washed three times in the same buffer, subjected to SDS-PAGE and visualized by autoradiography.

#### Plasmid construction

CAT gene reporter plasmids containing the E1b core promoter with wild-type (U4E1bCAT) and mutant (U4mE1bCAT) E-boxes, CAT gene reporter plasmids containing the AdML core promoter with wild-type (MLICAT) and mutated (MLIIR-I2CAT) initiator elements and a USF1 expression vector driven by the human cytomegalovirus immediate early promoter (pCX-USF1) were described previously (Du *et al.*, 1993). Expression vectors for the USF1 deletion constructs  $\Delta$ N39,  $\Delta$ N80,  $\Delta$ N93,  $\Delta$ N130,  $\Delta$ 94-130,  $\Delta$ 80-130,  $\Delta$ 58-130 and  $\Delta$ 26-130 were constructed by excising the *Nco*I-*Eco*RI fragments from the corresponding bacterial (pET3dUSF1) expression vectors (Kirschbaum *et al.*, 1992) and subcloning them in a shuttle vector, pBUSF1/*Nco*I-*Eco*RI, that provides the same human  $\beta$ -globin 5' untranslated region that is present in pCX-USF1. Then the *Xho*I-*Xba*I fragment of each mutant was transferred to the mammalian expression vector pCX. Other truncated USF1 expression plasmids were generated by PCR. pBUSF1 was used as template in PCR, and unique restriction sites (*Nco*I, *Avr*II and *Bgl*III) were used for cloning PCR-generated USF1 deletion fragments into pBUSF1 (pBUSF1/*Nco*I-*Bgl*III, pBUSF1/*Xba*I-*Bgl*III and pBUSF1/*Not*I-*Avr*II). The expression plasmid (pCX-II-I) for TFII-I was constructed by cloning the *Xba*I-*Eco*RI fragment of the full-length TFII-I cDNA (clone 3.1) into pCX digested with *Xba*I and *Eco*RI.

N-terminal deletion constructs  $\Delta$ N25,  $\Delta$ N100,  $\Delta$ N140 and  $\Delta$ N190 were generated by PCR using the primers:  $\Delta$ N25, 5'-GGCCATGGG-TGGGGAAAGACCCAACCAAGT;  $\Delta$ N100, 5'-GGCCATGGCCAG-TGATGATGCAGGTTGACACG;  $\Delta$ N140, 5'-GGCCATGGCCACTTACCAGGGCTCAGAGGGCA;  $\Delta$ N190, 5'-GGCCATGGCCGCTCCCGGACGAACTCGGGAT; and R1, 5'-CCGTTAAGATCTTCCACCT-GTTGTTCG. Forward primers containing a *Nco*I site (underlined) were annealed to pBUSF1 cDNA sequences at positions corresponding to amino acids 26, 101, 141 and 191, respectively. A reverse primer, R1, containing a *Bgl*III site (underlined) was annealed to pBUSF1 cDNA sequences at positions corresponding to amino acids 279-287.

For USF1 C-terminal deletion constructs  $\Delta$ 261-282,  $\Delta$ 231-282 and  $\Delta$ 216-282, the forward primer, L1A, was annealed to the upstream polylinker site of pBUSF1. Reverse primers: L1A, 5'-GTGGCCGC-CGCTCTAGAGTTCGACCTG;  $\Delta$ 261-282, 5'-GCCAGATCTGCGGT-GGTTACTCTGCCGAAG;  $\Delta$ 231-282, 5'-GCCAGATCTAGAGCA-GTCTGGGATTATCTT; and  $\Delta$ 216-282, 5'-GCCAGATCTCTGTCT-CTGGCCGCGACGCTC containing a *Bgl*III site (underlined) were annealed to pBUSF1 cDNA sequences at positions corresponding to amino acids 260, 230 and 215, respectively.

For USF1 internal deletion constructs  $\Delta$ 161-180,  $\Delta$ 131-180,  $\Delta$ 101-180,  $\Delta$ 81-180,  $\Delta$ 41-180 and  $\Delta$ 26-180, the forward primer, L1A, was annealed to the upstream polylinker site of pBUSF1. Reverse primers:  $\Delta$ 161-180, 5'-CTACCTAGGGAATTGACCAGTGCCAGGAGG;  $\Delta$ 131-180, 5'-CTACCTAGGACCCCTGCCCATCTCCAC;  $\Delta$ 101-180, 5'-CTACCTAGGGGTGAAAGCACCTGGATCAC;  $\Delta$ 81-180, 5'-CTAC-CTAGGGGCGCCAGTTCCTCAGTTT;  $\Delta$ 41-180, 5'-CTACCTA-GGGCAGCTGACTGGATGCTGGC; and  $\Delta$ 26-180, 5'-CTACCTA-GGCCAGTAGCCACTGCACCTTC containing an *Avr*II site (underlined) were annealed to pBUSF1 cDNA sequences at positions corresponding to amino acids 160, 130, 100, 80, 40 and 25.

PCRs containing 1 ng of pBUSF1 as a template and 100 pmol each of forward and reverse primers were incubated under conditions recommended by the manufacturer in a DNA thermal cycler (Perkin Elmer Cetus). After 30 cycles, 5% of the total PCR product was checked on a 1.2% agarose gel and aliquots of PCR products were end-filled by Klenow, digested with restriction enzymes corresponding to the sites designed in the primers and subcloned into the pBUSF1 vector. N-terminal deletions were ligated with pBUSF1/*Nco*I plus *Bgl*III, C-terminal deletions were ligated with pBUSF1/*Xba*I plus *Bgl*III, and internal deletions were ligated with pBUSF1/*Not*I plus *Avr*II. Each deletion fragment of USF1 in pBUSF1 was digested with *Xba*I and *Xho*I and subcloned into the expression vector pCX/*Xba*I plus *Xho*I. Each deletion construct was verified by dideoxynucleotide sequencing.

#### Transfection and CAT assays

*In vivo* transfections into HeLa cells and CAT assays were carried out as described by Chen and Okayama (1987) for the analysis in Figure 4B, and as described previously (Du *et al.*, 1993) for the analyses in Figures 4A, 5, 6B, 6C and 7. The difference is that transfected cells were incubated at 3% CO<sub>2</sub> in the former method, resulting in a slightly higher transfection efficiency, and at 5% CO<sub>2</sub> in the latter method. In all cases, the plasmid pGL2C (a luciferase gene driven by SV40 early promoter, Promega) was co-transfected as a reference for transfection efficiency. The transfected cell pellets were resuspended in 0.1 ml of 0.1 M KPO<sub>4</sub> (pH 8.2) and extracted by three freeze-thaw cycles at -70 and 37°C. The transfection efficiency was determined by luciferase assays using a photon counting luminometer. Adjusted amounts of extracts normalized to luciferase activity were used for CAT assays. Each transfection was repeated at least two to four times with different plasmid preparations. After autoradiography of the separated acetylated chloramphenicol forms, spots were excised and quantitated by liquid scintillation counting.

#### Accession number

The Genbank accession number for the TFII-I sequence is AF015553.

#### Acknowledgements

Microsequencing of TFII-I was performed at the Rockefeller microsequencing facility. We are grateful to Cindy Carruthers, Thomas Gutjahr, Bernhard Kirschbaum and other colleagues at the Rockefeller University for their contributions toward this project. We are also grateful to Dorre Grueneberg and Mike Gilman for sharing unpublished data. A.L.R. was supported by a postdoctoral fellowship from the Damon Runyon-Walter Wintchell Cancer Research Fund, P.D.G. by fellowships from the NIH (F32 AI07696) and the Lita A.Hazen Foundation, E.M. by fellowships from the Swiss National Science Foundation and the Revson-Winston Foundation and H.D. by an NIH training grant (CA09673). This work was funded in part by grants from the NIH (CA42567 and AI27397) to R.G.R., and by a grant from the Concern Foundation for Cancer Research to A.L.R.

#### References

Breathnach, R and Chambon, P. (1981) Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.*, **50**, 349-383.

- Conaway,R.C. and Conaway,J.W. (1993) General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.*, **62**, 161–190.
- Chen,C. and Okayama,H. (1987) High efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.*, **7**, 2745–2752.
- Chiang,C.M. and Roeder,R.G. (1995) Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. *Science*, **267**, 531–536.
- Dignam,J.D., Martin,P., Shastry,B. and Roeder,R.G. (1983) Eukaryotic gene transcription with purified components. *Methods Enzymol.*, **101**, 583–598.
- Du,H., Roy,A.L. and Roeder,R.G. (1993) Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the AdML promoters. *EMBO J.*, **12**, 501–511.
- Ferre-D'Amare,A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, **363**, 38–45.
- Ferre-D'Amare,A.R., Pognonec,P., Roeder,R.G. and Burley,S.K. (1994) Structure and function of the b/HLH/Z domain of USF. *EMBO J.*, **13**, 180–189.
- Gregor,P.D., Sawadogo,M. and Roeder,R.G. (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.
- Gruenberg,D.A., Henry,R.W., Brauer,A., Novina,C.D., Cheriya,V., Roy,A.L. and Gilman,M. (1997) A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: identity to TFII-I. *Genes Dev.*, **11**, 2482–2493.
- Hu,Y.F., Luscher,B., Admon,A., Mermod,N. and Tjian,R. (1990) Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Dev.*, **4**, 1741–1752.
- Javahery,R., Khachi,A., Lo,K., Zenzie-Gregory,B. and Smale,S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.*, **14**, 116–127.
- Kaufman,J. and Smale,S.T. (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.*, **8**, 821–829.
- Kaufman,J., Verrijzer,C.P., Shao,J. and Smale,S.T. (1996) CIF, an essential cofactor for TFIID-dependent initiator function. *Genes Dev.*, **10**, 873–886.
- Kirschbaum,B., Pognonec,P. and Roeder,R.G. (1992) Definition of the transcriptional activation domain of recombinant 43-kilodalton USF. *Mol. Cell. Biol.*, **12**, 5094–5101.
- Lathe,R. (1985) Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. *J. Mol. Biol.*, **183**, 1–12.
- Lu,H., Reach,M.D., Minaya,E. and Young,C.S.H. (1997) The initiator element of the adenovirus major late promoter has an important role in transcription initiation *in vivo*. *J. Virol.*, **71**, 102–109.
- Luo,X. and Sawadogo,M. (1996) Functional domains of the transcription factor USF2: atypical nuclear localization signals and context-dependent transcriptional activation domains. *Mol. Cell. Biol.*, **16**, 1367–1375.
- Manzano-Winkler,B., Novina,C.D. and Roy,A.L. (1996) TFII-I is required for transcription of the naturally TATA-less but initiator-containing V $\beta$  promoter. *J. Biol. Chem.*, **271**, 12076–12081.
- Martinez,E., Chiang,C.M., Ge,H. and Roeder,R.G. (1994) TAFs in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.*, **13**, 3115–3126.
- Martinez,E., Zhou,Q., L'Etoile,N.D., Oelgeschlager,T., Berk,A.J. and Roeder,R.G. (1995) Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding. *Proc. Natl Acad. Sci. USA*, **92**, 11864–11868.
- Meisterernst,M. and Roeder,R.G. (1991) Family of proteins that interact with TFIID and regulate promoter activity. *Cell*, **67**, 557–567.
- Molkentin,J.D., Black,B.L., Martin,J.F. and Olson,E.N. (1995) Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell*, **83**, 1125–1136.
- Molkentin,J.D. and Olson,E.N. (1996) Combinatorial control of muscle development by basic helix–loop–helix and MADS-box transcription factors. *Proc. Natl Acad. Sci. USA*, **93**, 9366–9373.
- Murre,C. and Baltimore,D. (1992) *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Novina,C.D. and Roy,A.L. (1996) Core promoter mediated transcription. *Trends Genet.*, **12**, 351–355.
- Orphanides,G., Lagrange,T. and Reinberg,D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev.*, **10**, 2657–2683.
- Outram,S.V. and Owen,M.J. (1994) The helix–loop–helix containing transcription factor USF activates the promoter of the CD2 gene. *J. Biol. Chem.*, **269**, 26525–26530.
- Pognonec,P. and Roeder,R.G. (1991) Recombinant 43-kDa USF binds to DNA and activates transcription in a manner indistinguishable from that of 43/440kDa USF. *Mol. Cell. Biol.*, **11**, 5125–5136.
- Ptashne,M. and Gann,A. (1997) Transcriptional activation by recruitment. *Nature*, **386**, 569–577.
- Roeder,R.G. (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem. Sci.*, **16**, 402–408.
- Roeder,R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.*, **21**, 327–335.
- Roy,A.L. and Roeder,R.G. (1994) Initiator element binding protein TFII-I: a tale of two sites. *Ind. J. Biochem. Biophys.*, **31**, 14–19.
- Roy,A.L., Meisterernst,M., Pognonec,P. and Roeder,R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix–loop–helix activator USF. *Nature*, **354**, 245–248.
- Roy,A.L., Malik,S., Meisterernst,M. and Roeder,R.G. (1993a) An alternative pathway for transcription initiation involving TFII-I. *Nature*, **365**, 355–359.
- Roy,A.L., Carruthers,C., Gutjahr,T. and Roeder,R.G. (1993b) Direct role for MYC in transcription initiation mediated by interactions with TFII-I. *Nature*, **365**, 359–361.
- Sawadogo,M. and Roeder,R.G. (1995) Interaction of a gene specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell*, **43**, 165–175.
- Sawadogo,M., Van Dyke,M.W., Gregor,P.D. and Roeder,R.G. (1988) Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa nuclei. *J. Biol. Chem.*, **263**, 11985–11993.
- Scheidereit,C., Cromlish,J.A., Gerster,T., Kawakami,K., Balmaceda,C.-G., Currie,R.A. and Roeder,R.G. (1988) A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox protein. *Nature*, **336**, 551–557.
- Seto,E., Shi,Y. and Shenk,T. (1991) YY1 is an initiator sequence binding protein that directs and activates transcription *in vitro*. *Nature*, **354**, 241–245.
- Shore,P. and Sharrocks,D. (1995) The MADS-box family of transcription factors. *Eur. J. Biochem.*, **229**, 1–13.
- Smale,S.T. (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta*, **1351**, 73–88.
- Smale,S.T. and Baltimore,D. (1989) The 'initiator' as a transcription control element. *Cell*, **57**, 103–113.
- Triezenberg,S.J. (1995) Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.*, **5**, 190–195.
- Usheva,A. and Shenk,T. (1994) TATA binding protein independent initiation: YY1, transcription factor IIB and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell*, **76**, 1115–1121.
- Verrijzer,C.P. and Tjian,R. (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.*, **21**, 338–342.
- Verrijzer,C.P., Chen,J.L., Yokomori,K. and Tjian,R. (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**, 1115–1125.
- Weintraub,H., Genetta,T. and Kadesch,T. (1994) Tissue-specific gene activation by MyoD: determination of specificity by *cis*-acting repression elements. *Genes Dev.*, **8**, 2203–2211.
- Wiley,S.R., Kraus,R.J. and Mertz,J.E. (1992) Functional binding of the 'TATA' box binding component of transcription factor TFIID to the –30 region of TATA-less promoters. *Proc. Natl Acad. Sci. USA*, **89**, 5814–5818.
- Yang,W. and Desiderio,S. (1997) BAP-135, a target for Bruton's tyrosine kinase in response to B cell receptor engagement. *Proc. Natl Acad. Sci. USA*, **94**, 604–609.
- Zawel,L. and Reinberg,D. (1995) Common themes in assembly and function of eukaryotic transcriptional complexes. *Annu. Rev. Biochem.*, **64**, 533–561.

Received on July 9, 1997; revised on August 28, 1997