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

Polly D. Gregor, Michele Sawadogo,[1] and Robert G. Roeder

Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, New York 10021-6999 USA

We isolated full-length cDNAs encoding the 43-kD form of human upstream stimulatory factor (USF), a cellular factor required for efficient transcription of the adenovirus major late (AdML) promoter in vitro. Sequence analysis showed USF to be a member of the c-myc-related family of DNA-binding proteins. Using proteins translated in vitro, we identified a DNA-binding domain near the carboxyl terminus, which includes both a helix–loop–helix motif and a leucine repeat. We show that USF interacts with its target DNA as a dimer. The leucine repeat is required for efficient DNA binding of the intact protein and for interactions between full-length and truncated USF proteins. Interestingly, it is not required for DNA binding of the isolated helix–loop–helix domain. The structure of different cDNA clones indicates that USF RNA is differentially spliced, and alternative exon usage may regulate the levels of functional USF protein.

[Key Words: USF; c-myc-related DNA-binding proteins; helix–loop–helix]

Sequences within the adenovirus major late (AdML) promoter that are required for its efficient expression in vitro and in vivo have been identified by extensive mutational analysis in several laboratories (Hu and Manley 1981; Hen et al. 1982; Concino et al. 1984; Miyamoto et al. 1984; Yu and Manley 1984; Sawadogo and Roeder 1985; Logan and Shenk 1986; Brunet et al. 1987; Lee et al. 1988; Smale and Baltimore 1989). Altogether, these studies have identified three promoter elements; the upstream activating element (UAE) centered at position −63 to −52, the TATA box located at position −31 to −25, and the initiator at position −6 to +4.

Sawadogo and Roeder (1985) and others (Carthew et al. 1985; Miyamoto et al. 1985) have described a cellular factor, upstream stimulatory factor (USF, or MLTFI or UEF) that both stimulates transcription of the AdML promoter up to 10-fold in vitro and binds to an upstream element that is essential for this stimulation. Glycerol gradient sedimentation studies indicate a size of 45–60 kD for USF (Chodosh et al. 1986; Moncollin et al. 1986). Near homogeneous preparations of USF contain two polypeptides of 43 and 44 kD, and each form binds DNA specifically (Sawadogo et al. 1988). These preparations also stimulate in vitro transcription but less dramatically than the partially purified factor (Sawadogo 1988). DNase I and MPE · Fe (II) footprinting (Carthew et al.

1985; Sawadogo and Roeder 1985; Lennard and Egly 1987) and methylation interference (Carthew et al. 1987; Lennard and Egly 1987) studies have identified the site of USF binding as the nearly symmetrical sequence GGCCACGTGACC.

USF is present in diverse cell types (Carthew et al. 1987). DNA-binding and transcriptional studies have implicated USF in the expression of several tissue-specific or developmentally regulated genes, including human growth hormone (Peritz et al. 1988), mouse metallothionein I (Carthew et al. 1987), rat γ fibrinogen (Chodosh et al. 1987), and Xenopus TFIIIA (Scotto et al. 1989). In addition, USF interacts with an essential DNA motif (uE3) in the immunoglobulin heavy-chain enhancer (L. Poellinger et al., in prep.). This broad range of functions suggests that USF interacts with both general and cell-type-specific factors. Potentially relevant to its mechanism of action, cooperative interactions between USF and the TATA-binding factor TFIID were demonstrated on the ML promoter (Sawadogo and Roeder 1985; Sawadogo 1988; Workman et al. 1990).

To examine in detail the mechanisms by which USF activates transcription, we have isolated human USF-specific cDNAs. Here, we describe these clones and present a preliminary characterization of the USF DNA-binding domain. We show that USF is a member of a c-myc-related family of regulatory factors containing helix–loop–helix domains (Murre et al. 1989a) and that it binds to DNA as a dimer. USF also contains a leucine

repeat (Landschulz et al. 1988a) that is required for efficient DNA binding of the full-length protein, and for dimerization of both full-length and truncated proteins. Interestingly, a truncated protein containing the helix–loop–helix domain but lacking the leucine repeat binds DNA specifically but does not dimerize with other USF proteins. Thus, the leucine repeat has a complicated role in DNA recognition by USF. We propose that in the context of the full-length protein the leucine repeat is required to maintain the helix–loop–helix domain in a functional conformation and that it could also promote interactions between USF and other proteins in vivo.

Results

Isolation of USF cDNA clones

USF was purified to near homogeneity by conventional chromatography (Sawadogo et al. 1988), and the 43- and 44-kD subspecies were resolved by SDS–polyacrylamide gel electrophoresis (SDS-PAGE). One-hundred pico-moles of the 43-kD form of USF was transferred to Immobilon (Millipore), allowing the amino-terminal sequence to be obtained. Residues 1–25 were identified, with positions 2–14 unambiguous (Materials and methods). In a parallel experiment, 30 pmoles of the 44-kD USF protein yielded no sequence, suggesting the presence of a blocked amino terminus. A synthetic peptide containing residues 2–14 of the 43-kD form of USF was used to raise a rabbit antisera that showed specificity to USF both in Western blots and by interaction with USF–DNA complexes in a mobility-shift assay (data not shown). This analysis provided confirming evidence that the experimentally determined amino acid sequence was indeed from USF.

For cloning, the sequence of residues 1–14 was used to synthesize a 42-nucleotide "best guess" oligonucleotide (Lathe 1985), which was used to screen a human Namalwa cell cDNA library cloned in λ ZAP by Stratagene. Eight phage clones were plaque-purified from 17 primary positives detected in a screen of 2×10^6 phage. Sequence analysis of the excised inserts yielded three independent clones, z6b, z18, and z32, which contained sequences encoding amino acids 1–25 in USF, in complete agreement with the data obtained from microsequencing of the purified protein (Materials and methods).

USF mRNAs are differentially spliced

Clones z18 and z32 were sequenced in their entirety, and z6b was characterized by restriction enzyme mapping and partial sequencing. All three clones are assumed to be full length, as they each contain the amino-terminal USF sequence at the start of an open reading frame and terminate in a poly(A) tail. Clones z18 and z32 diverge in the 5'-untranslated region 50 bp upstream of the initiating AUG, and the natural start site for transcription has not been determined. The 3'-untranslated region is identical in z18 and z32, differing only in the length of the poly(A) tail. A potential polyadenylation

signal sequence (GATAAA) is located 19 bp 5' of this poly(A) tail and is preceded by a stretch of 27 Ts.

As summarized in Figure 1, clone z18 includes two sequences of 133 and 245 bp that are not found in z32. These sequence elements could be exons that are sometimes excised by differential RNA splicing. A genomic clone is required to confirm this hypothesis, and here these regions are referred to simply as optional domains 1 and 2 (OD1 and OD2). OD1 contains a single continuous open reading frame. Its absence in z32 results in a frameshift that introduces a stop codon after 2 residues, thus predicting a 49-amino-acid polypeptide. Clone z18 has an open reading frame of 244 amino acids, terminating 6 amino acids within OD2. Clone z6b contains OD1, but not OD2, and has an open reading frame of 310 amino acids, encoding a protein of 33,517 daltons. This open reading frame includes three blocks of c-*myc*-related amino acids: the *myc* similarity regions one and two (MSR1 and MSR2) and a leucine repeat (LR). This sequence places USF in the c-*myc*-related family of nuclear proteins, which includes the enhancer-binding proteins E12, E47 (Murre et al. 1989a), and TFE3 (Beckmann et al. 1990) and cell determination factors *achaete-scute* (Villares and Cabrera 1987), *daughterless* (Caudy et al. 1988), MyoD (Davis et al. 1987), and myogenin (Wright et al. 1989). Interestingly, MSR1 and MSR2 are separated by OD2 in z18 (Fig. 1).

Polymerase chain reaction (PCR) analysis was used to identify USF cDNAs in the Namalwa library. By using oligonucleotides complementary to sequences 5' of OD1 and 3' of OD2, PCR products identical in size to those predicted from the z6b, z18, and z32 sequences were detected when the entire cDNA library was used as substrate (results not shown). To determine the frequency of each class of cDNA, a panel of 54 primary positives, isolated by homology either to the 42-mer best guess probe or the entire z18 insert, was screened by using PCR. Initially, pools of nine clones were tested. Members of pools that directed the synthesis of discrete PCR products were assayed individually. These positive clones included two identical to z18, five identical to z6b, and none identical to z32 (data not shown).

In an electrophoretic mobility-shift assay, neither of the proteins encoded by z18 and z32 retarded an oligonucleotide probe containing the USF-binding site from the AdML promoter (data not shown). In our initial analysis, the *EcoRI–KpnI* fragment from z18 was ligated to the *KpnI–EcoRI* fragment of z32 to create a clone (dI2) with the splice pattern of z6b. The protein produced by in vitro transcription and translation of either z6b or dI2 has an apparent molecular mass of 43 kD by SDS-PAGE (Fig. 4C) and binds DNA in a sequence-specific manner (e.g., see Fig. 4A, lane 2). The nucleotide and deduced amino acid sequence of dI2 is presented in Figure 2. Because the coding regions of clones z6b and dI2 appear to be identical, we used them interchangeably.

As described above, we cloned the 43-kD form of USF on the basis of amino acid sequence obtained from a purified, transcriptionally active protein. To obtain further evidence that our cDNA encodes a transcriptionally ac-
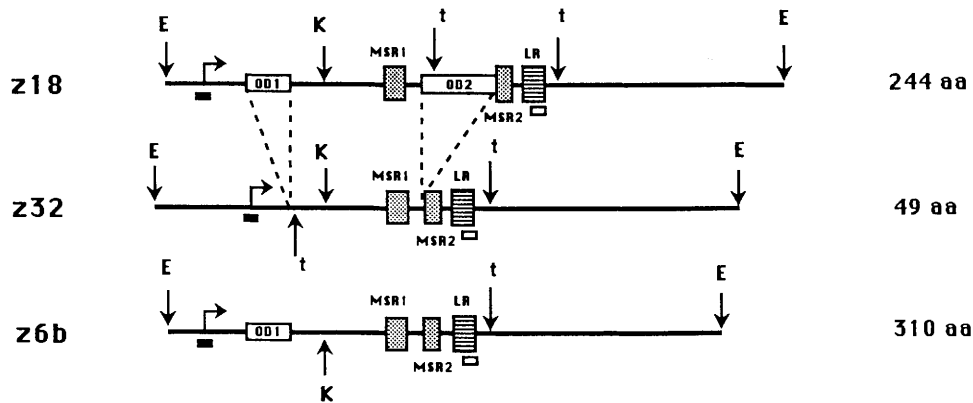
Figure 1. cDNAs encoding USF. Line drawing of *Eco*RI inserts. OD1 and OD2 are indicated by open boxes, MSR1 and MSR2 are indicated by stippled boxes, and the leucine repeat is indicated by a striped box. Bent arrows are located at initiating methionines, and straight vertical arrows (t) are located at translation termination sites. Protein sizes predicted from the DNA sequence are at *right*. The locations of oligonucleotides used for PCR analysis are indicated below the line by solid (Nco) and open (LR) boxes. Restriction sites are *Eco*RI (E) and *Kpn*I (K).

tive protein, it was expressed in bacteria, and the resulting protein was purified and assayed in a system reconstituted with purified general transcription factors. The transcriptional activity of the bacterially expressed protein (M. Meisterernst, P. Pognonec, P.D. Gregor, and R.G. Roeder, unpubl.) was equivalent to that observed for the highly purified HeLa USF (Sawadogo 1988); however, both USF sources stimulated weakly when compared to a crude HeLa USF preparation, suggesting the involvement of an additional cofactor not present in the purified reconstituted system.

The involvement of the 43-kD protein in transcription was also demonstrated by using USF-specific antisera raised in rabbits against the bacterially expressed protein. Specific immunodepletion of HeLa nuclear extracts reduced transcription of a wild-type AdML promoter that contained an intact UAE but had no effect on transcription of the same promoter truncated at position − 53. Immunodepletion with naive serum had no effect on transcription from either promoter (P. Pognonec and R.G. Roeder, unpubl.). The reduction in USF-dependent transcription was comparable to that seen for nuclear extracts depleted using a USF-specific oligonucleotide–sepharose affinity matrix. Together, these results clearly indicate that the 43-kD form of USF is an essential component of the previously described USF activity but do not exclude the possibility that other components may contribute to that activity.

Amino acid sequence of dl2

As noted above, USF contains a region of 99 amino acids with significant sequence similarity to c-*myc*. This region can be subdivided into a basic region, two potential amphipathic helices separated by a putative "loop" (Murre et al. 1989a), and a leucine repeat (Landschulz et al. 1988a). In Figure 3, the amino acid sequence of USF is compared to other c-*myc*-related proteins. Within the putative amphipathic helices and the leucine repeat, similarities between family members are restricted to

the conserved residues (shaded) that are believed to confer structure to these domains. In no case does the sequence homology between USF and other family members extend beyond the *myc* similarity regions (comprising the basic region/helix 1 and helix 2 regions) and the leucine repeat motifs.

Acidic residues are distributed evenly in USF, with no clusters like those identified in the transcriptional activation domains of other proteins (Ma and Ptashne 1987; Hope et al. 1988). However, computer projections predict that the amino-terminal 22 residues can adopt an amphipathic helical structure with charged or polar residues at positions 1, 3, and 4 in the helix. This region may be important for transcriptional activity, as it resembles the structure, although not the sequence, of a transcriptional activation domain in TFE3 (Beckmann et al. 1990).

Regions rich either in glutamine residues or in serine and threonine residues are important in transcriptional activation by SP1 (Mitchell and Tjian 1989) and OTF2 (Gerster et al. 1990; Muller-Immergluck et al. 1990; Tanaka and Herr 1990). USF contains 29 glutamines (9.4%) that are distributed evenly throughout the protein. However, 4 glutamine residues within the putative α-helix at the amino terminus contribute to its amphipathic nature. In addition, we noticed that of the 15 glutamine residues within the first 160 amino acids of USF, 6 are directly preceded by a glycine, 4 by an isoleucine, and 3 by a threonine residue. There are no obvious pairings of amino acids with glutamine elsewhere in the protein. Serines and threonines are evenly distributed in USF, except for the virtual absence of threonine (1 of 28 residues) in the *myc* similarity regions and the leucine repeat and a cluster of serines and threonines (5 of 6 residues) from 131–136. There is also a cluster of serines (5 of 17 residues) in the putative loop between MSR1 and MSR2.

A repeated sequence S–S-APRT—S–S-APRT lies between residues 174 and 195. This short repeat resembles sequences in the hinge region of immunoglobulin genes

```
GGAATTCCTT  GAAAATTTTC  CTTGGATAGG  AAAGGACTTA  GCACTCAGGC  CTGTGAATCT
AGGAGATACA  AAGACCTCCA  AAAAAGGACC  AGTTCCTCGG  ATGTGCCCCC  TCACAGAGAG

     ATG AAG GGG CAG CAG AAA ACA GCT GAA ACG GAA GAG GGG ACA GTG CAG ATT CAG GAA GGT        180
  1  M   K   G   Q   Q   K   T   A   E   T   E   E   G   T   V   Q   I   Q   E   G

     GCA GTA GCT ACT GGG GAA GAC CCA ACA AGT GTG GCT ATT GCC AGC ATC CAA TCA GCT GCC        240
 21  A   V   A   T   G   E   D   P   T   S   V   A   I   A   S   I   Q   S   A   A

     ACC TTC CCT GAC CCC AAC GTC AAG TAC GTC TTC CGA ACT GAG AAT GGG GGC CAG GTG ATG        300
 41  T   F   P   D   P   N   V   K   Y   V   F   R   T   E   N   G   G   Q   V   M

     TAC AGG GTG ATC CAG GTG TCT GAG GGG CAG CTG GAT GGC CAA ACT GAG GGA ACT GGC GCC        360
 61  Y   R   V   I   Q   V   S   E   G   Q   L   D   G   Q   T   E   G   T   G   A

     ATC AGT GGC TAC CCT GCC ACT CAA TCC ATG ACC CAG GCG GTG ATC CAG GGT GCT TTC ACC        420
 81  I   S   G   Y   P   A   T   Q   S   M   T   Q   A   V   I   Q   G   A   F   T

     AGT GAT GAT GCA GTT GAC ACG GAG GGG ACA GCT GCT GAA GCA CAC TAT ACT TAC TTC CCC        480
101  S   D   D   A   V   D   T   E   G   T   A   A   E   T   H   Y   T   Y   F   P

     AGC ACG GCA GTG GGA GAT GGG GCA GGG GGT ACC ACA TCG GGG AGT ACA GCT GCT GTT GTT        540
121  S   T   A   V   G   D   G   A   G   G   T   T   S   G   S   T   A   A   V   V

     ACT ACC CAG GGC TCA GAG GCA CTG CTG GGG CAG GCA ACC CCT CCT GGC ACT GGT CAA TTC        600
141  T   T   Q   G   S   E   A   L   L   G   Q   A   T   P   P   G   T   G   Q   F

     TTT GTG ATG ATG TCA CCA CAA GAA GTA CTG CAG GGA GGA AGC CAG CGC TCA ATT GCC CCT        660
161  F   V   M   M   S   P   Q   E   V   L   Q   G   G   S   Q   R   S   I   A   P
                                                                         181
     AGG ACC CAT CCT TAT TCC CCG AAG TCA GAA GCT CCC CGG ACG ACT CGG GAT GGA AAA CGC        720
181  R   T   H   P   Y   S   P   K   S   E   A   P   R   T   T   R   D   E   K   R

     AGG GCT CAG CAT AAT GAA GTG GAG CGT CGC CGC CGA GAC AAG ATC AAC AAC TGG ATC GTG        780
201  R   A   Q   H   N   E   V   E   R   R   R   R   D   K   I   N   N   W   I   V

     CAG CTC TCC AAG ATA ATC CCA GAC TGC TCT ATG GAG AGC ACC AAG TCT GGC CAG AGT AAA        840
221  Q   L   S   K   I   I   P   D   C   S   M   E   S   T   K   S   G   Q   S   K

     GGT GGG ATT CTA TCC AAA GCT TGT GAT TAT ATC CAG GAG CTT CGG CAG AGT AAC CAC CGC        900
241  G   G   I   L   S   K   A   C   D   Y   I   Q   E   L   R   Q   S   N   H   R

     TTG TCT GAA GAA CTG CAG GGA CTT GAC CAA CTG CAG CTG GAC AAT GAC GTG CTT CGA CAA        960
261  L   S   E   E   L   Q   G   L   D   Q   L   Q   L   D   N   D   V   L   R   Q
                          284                    292
     CAG GTG GAA GAT CTT AAA AAC AAG AAT CTG CTG CTT CGA GCT CAG CTG CGG CAC CAC GGA       1020
281  Q   V   E   D   L   K   N   K   N   L   L   L   R   A   Q   L   R   H   H   G
                                             310
     TTA GAG GTC GTC ATC AAG AAT GAC AGC AAC TAA                                           1053
301  L   E   V   V   I   K   N   D   S   N   End
```

```
C⁻ATGGGGAT  TCAGGGGCTT  TGGGCCCAAG  AACTGCAGAT  AGCCCAGGAG  CAACAGCCTA
ATCCCGTGCC  CC⁻TTCCTTC  ACTGCCCCAC  TTCTGGCATG  GGACAGGGGG  AAGTTCAGAA
GGTGTGTCCT  TGAACTGACG  CCCTGTGATA  TGGCGGCCTG  CAGTGGTGTG  AAGCACACAA
TGTGGAAGCT  GCACTGACAG  CCTTGCCCAC  CCCCACCATG  CAGCCCCTGG  GCCTTGTGCT
CCTCTCGCAC  AATGCATGTG  CTGTCTCCAT  GCTGGATACT  GGACACACTA  AACTCTGGGG
CTTGTCCTGT  GCTTGCTTAG  AGTGCCCAGC  AGAGGTTTGC  TGACAGGTGA  TGCTCTGGCT
TGCCCCAGGA  CTCTGGCACT  TCCATTGGTT  CTTCCTTTCC  CTGGAGCTGA  GGTTTAGATG
TGCAACCTGT  GGCTCAGGGG  AGCAAGCTTA  CACAAGAAGT  GAGGGAAGGA  TGTTTAGCAG
TGGCTGGTGC  CCATGAAGAG  GAGATTGGCC  AGTGAGAAGC  TGAGGCCTAT  GCAGACATCT
CTGGAGCCAG  AGAGAACAAC  AGGCAGGGGC  CCACTTGGGG  CCTTCCCCCT  TGTGGGGGGT
CGTTTTTTTT  TTTTCTTTTC  TTTTTTTTTT  TTTTTTTTTT  TTTTTTTAAG  ATAAAATTGT
TCAAAGCCAA  AAAAAAAAGG  AATTCC
```
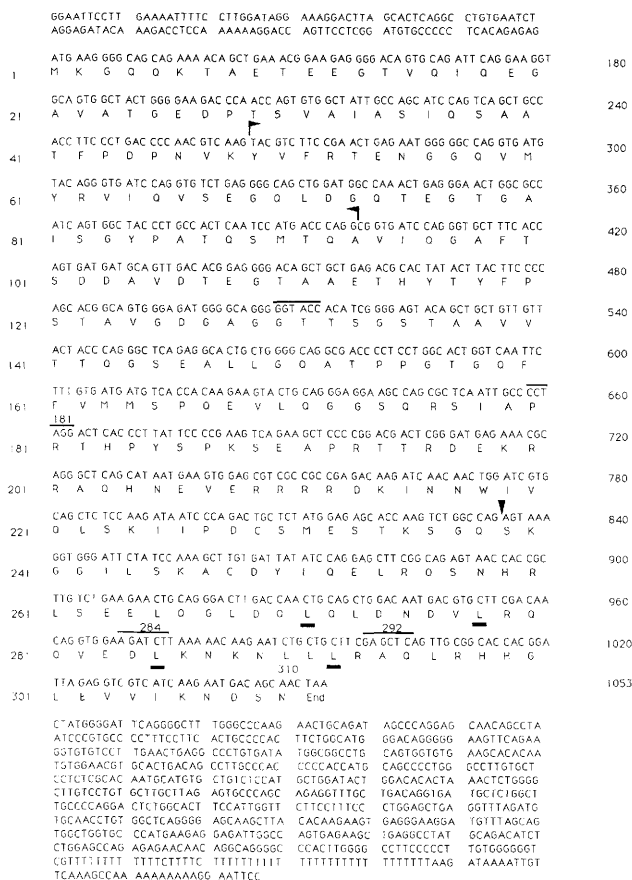
Figure 2. Nucleotide and deduced protein sequence of the cDNA dI2. Numbers to the *left* of the sequence indicate amino acids; those to the *right* indicate nucleotides. The 5'-untranslated region and the entire coding region were sequenced in both directions. Portions of the 3'-untranslated region were sequenced on only one strand. The left and right boundaries of OD1 are indicated by flags. The location of excised OD2 is indicated by a solid inverted triangle. Solid lines above the sequence locate restriction sites used to subclone or truncate templates (see text), and the accompanying numbers give the expected size of the resulting protein. The nucleotide sequence data reported has been submitted to the EMBL/GenBank and DDBJ Nucleotide Sequence Databases.

and could serve an analogous function in USF by separating DNA-binding and transcriptional activation domains. A second repeated peptide sequence, TEGT-A-S-Y../..TEGT-A-T-Y is located between residues 74 and 116.

Efficient DNA binding requires an intact leucine repeat

To map the sequences required for DNA binding, cDNA templates were cleaved at restriction sites in the 3' polylinker or within the coding region and transcribed with T7 RNA polymerase. The transcripts were translated in a rabbit reticulocyte lysate to produce full-length or carboxy-terminal-truncated proteins (Fig. 4C). The full-length protein of 310 amino acids bound a radiolabeled oligonucleotide (ML; see Materials and methods) containing sequences corresponding to posi-
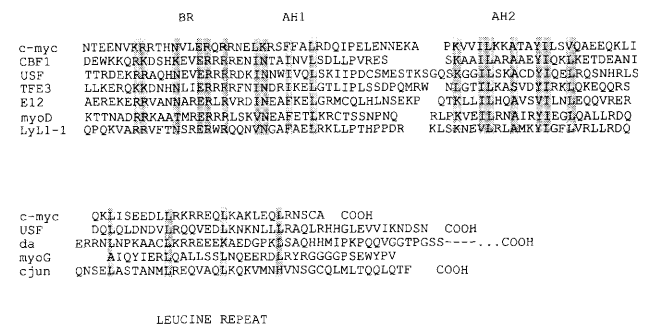
```
              BR              AH1                              AH2
c-myc   NTEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKA    PKVVILKKATAYILSVQAEEQKLI
CBF1    DEWKKQRKDSHKEVERRRENINTAINVLSDLLPVRES         SKAAILARAAEYIQKLKETDEANI
USF     TTRDEKRRAQHNEVERRRRDKINNWIVQLSKIIPDCSMESTKSGQSKGGILSKACDYIQELRQSNHRLS
TFE3    LLKERQKKDNHNLIERRRRFNINDRIKELGSTLIPLSSDPQMRW  NLGTILKASVDYIRKLQKEQQRS
E12     AEREKERRVANNARERLRVRDINEAFKELGRMCQLHLNSEKP    QTKLLILHQAVSVILNLEQQVRER
myoD    KTTNADRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQ       RLPKVEILRNAIRYIEGLQALLRDQ
LyL1-1  QPQKVARRVFTNSRERWRQQNVNGAFAELRKLLPTHPPDR      KLSKNEVLRLAMKYIGFLVRLLRDQ
```

```
c-myc   QKLISEEDLLRKRREQLKAKLEQLRNSCA    COOH
USF     DQLQLDNDVLRQQVEDLKNKNLLLRAQLRHHGLEVVIKNDSN    COOH
da      ERRNLNPKAACLKRREEEKAEDGPKDSAQHHMIPKPQQVGGTPGSS----...COOH
myoG    AIQYIERLQALLSSLNQEERDLRYRGGGGPSEWYPV
cjun    QNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF     COOH
```

LEUCINE REPEAT

Figure 3. Sequence alignment in the MSR of USF. Sequences are from human c-*myc* (amino acids 341–429; Battey et al. 1983), CBF1 (amino acids 217–278; Cai and Davis 1990), TFE3 (amino acids 134–199; Beckmann et al. 1990), E12 (amino acids 331–395; Murre et al. 1989a), MyoD (amino acids 104–167; Davis et al. 1987), Lyl-1 (amino acids 132–196; Mellentin et al. 1989), *daughterless* (amino acids 613–657; Caudy et al. 1988), myogenin (amino acids 123–158; Wright et al. 1989), and c-Jun (amino acids 299–340; Bohmann et al. 1987). The basic region (BR), proposed amphipathic helices 1 and 2 (AH1 and AH2) (Murre et al. 1989a), and leucine repeat are indicated. The most conserved residues are shaded.

tions −75 to −45 of the AdML promoter, and binding was specifically inhibited by an excess of the corresponding unlabeled ML oligonucleotide (Fig. 4A, lanes 1 and 2). The major DNA–protein complex comigrated with the larger of two complexes observed with natural HeLa-derived USF (lane 3). A control lane with lysate programmed with Brome mosaic virus (BMV) RNA showed a small amount of specific complex that represents endogenous rabbit USF. This complex was both competed by the ML oligonucleotide (Fig. 4A, lane 1; Fig. 7, lanes 2–4 and 9–11) and recognized by a mouse polyclonal anti-HeLa USF sera (results not shown).

To define the carboxy-terminal residues required for efficient DNA binding, USF cDNA templates were digested with restriction enzymes, and truncated proteins were assayed for DNA binding (Materials and methods). A protein containing amino acids 1–292 bound DNA less strongly than the wild-type protein and gave a fainter, more diffuse band (Fig. 4B, lane 2). This may result from an instability of these DNA–protein complexes. A protein containing amino acids 1–284, including only half of the leucine repeat, showed a further reduction in binding to a level equivalent to that observed with the BMV control RNA (Fig. 4B, lanes 3 and 4). Increasing the amount of the 1- to 284-amino acid protein in the assay resulted in a small but significant increase in DNA binding relative to a BMV control (results not shown). From this, we conclude that the leucine repeat and possibly other carboxy-terminal residues are required for efficient DNA binding.

USF binds to DNA as a dimer

DNA-binding proteins that recognize palindromic DNA sequences commonly bind DNA as dimers (Pabo and Sauer 1984). However, analysis by scanning transmission electron microscopy (Hough et al. 1987) and kinetic
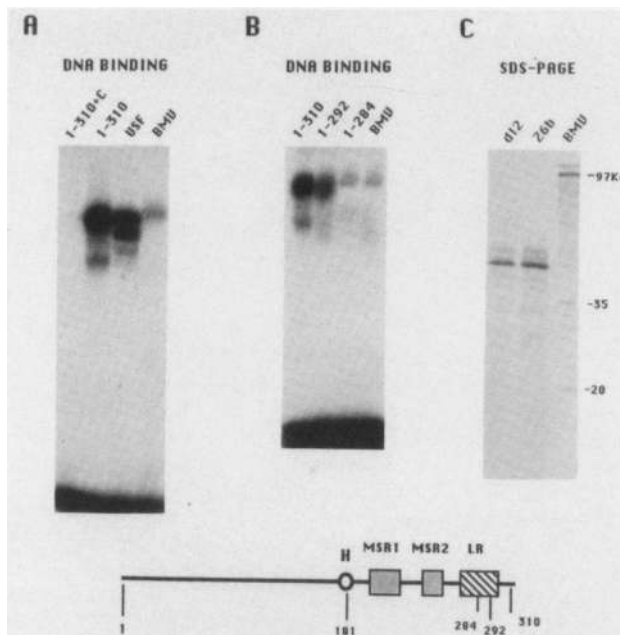
Figure 4. Sequence-specific binding of the dI2 protein requires the leucine repeat domain. (A) Mobility shift assays, using the ML oligonucleotide as probe. Protein sources were (1–310) reticulocyte lysate programmed with the full-length cDNA; (USF) partially purified human USF; (BMV) reticulocyte lysate programmed with Brome mosaic virus RNA. One-hundred nanograms of unlabeled ML oligonucleotide was added to one reaction (1-310 + C). HeLa USF gives a doublet; the intensity of the lower band varies from preparation to preparation and may therefore arise from proteolysis. The minor band (of higher mobility) in the USF programmed lysate may result from initiation of an internal methionine. (B) Mobility-shift assays, using the ML oligonucleotide as probe. Proteins were produced by sequential in vitro transcription and translation by using templates linearized with restriction enzymes XhoI (amino acids 1–310), SacI (amino acids 1–292), or BglII (amino acids 1–284). The map (bottom) indicates the sites of template truncation, and the location of the hinge (H), MSR1, MSR2, and leucine repeat domains (LR) (see text). (C) SDS-PAGE of in vitro translation products. Proteins (5-μl lysate) translated from z6b or dI2 templates were separated by SDS-PAGE and transferred to nitrocellulose. Molecular weight markers were provided by 1 μl of BMV-programmed lysate. The labeled USF proteins comigrate exactly with USF isolated from HeLa cells and electrophoresed in parallel, as seen in Western blots stained with the rabbit antipeptide antiserum (data not shown).

analysis of USF–DNA interactions (Lennard and Egly 1987) indicated that USF binds DNA as a monomer. The amino acid sequence of the cloned protein predicts a monomer of 33.5 kD or a dimer of 67 kD. Neither possibility is precluded by the previously determined mass of USF in solution (45–60 kD). To determine experimentally whether our recombinant USF binds DNA as a monomer or as a dimer, wild-type (1–310) and amino-terminally truncated (181–310) USF proteins were translated together or separately and assayed for DNA binding (Fig. 5). The truncated protein 181–310, protein B in Figure 5, apparently contains all residues required for sequence-specific DNA binding.

Upon cotranslation of proteins A (1–310) and B (181–310) a DNA–protein complex of intermediate mobility appeared, which we believe reflects a heteromer comprised of both proteins A and B. The presence of a single intermediate is most consistent with a dimeric structure for the heteromer, but the data would also be consistent with a tetramer if the interactions were subject to symmetry constraints (e.g., if only symmetrical dimers could form and interact). This heteromer was not seen when proteins derived from individually translated RNAs were simply mixed, indicating that heteromers, once formed, are extremely stable and that monomers do not exchange freely in solution. These data are consistent both with the prediction of a protein dimer(s) interacting with a symmetrical DNA-binding site and with the apparent size (45–60 kD) of the protein in solution (Chodosh et al. 1986; Moncollin et al. 1986) and bound to DNA (Hough et al. 1987; Lennard and Egly 1987). Given these observations and for purposes of discussion, we tentatively refer to the heteromer as a dimer.

The leucine repeat region is not required for site-specific binding of an amino-terminal-truncated USF

Because the leucine repeat is required for efficient DNA binding (Fig. 4) and because USF binds DNA as a dimer (Fig. 5), we tested the role of the leucine repeat in protein–protein interactions. Proteins containing residues 181–310 and proteins with additional truncations at the carboxyl terminus were cotranslated with the full-length 310 amino acid protein. The amount of full-
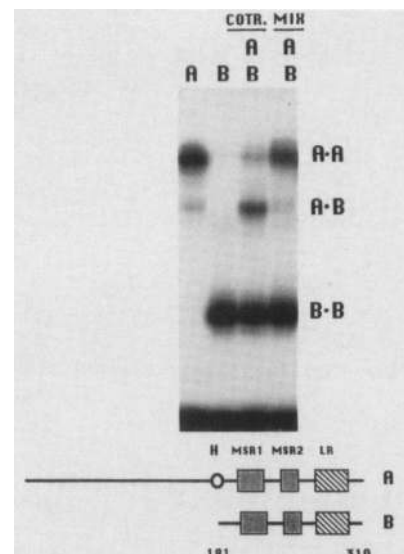

Figure 5. USF binds as a dimer. Mobility-shift assays, using the ML oligonucleotide as probe. Proteins A (amino acids 1–310) and B (amino acids 180–310) were translated in vitro together (COTR) or separately and then mixed in the DNA binding assay (MIX). Predicted dimers A · A, A · B, and B · B are indicated. Maps of the A and B proteins are shown (bottom) and indicate the location of the hinge (H), MSR1, MSR2, and leucine repeat (LR) domains (see text).

length (wt) RNA was kept constant in each cotranslation, and increasing amounts of truncated RNAs were added to the reticulocyte lysate. A heteromer between 1–310 (wt) and 181–310 (a) proteins appeared as a single species of intermediate mobility and increased in intensity as the 181–310 protein was increased (Fig. 6, left).

The effect of additional carboxy-terminal truncations on DNA binding of the amino-terminal-truncated protein containing residues 181–310 is shown in Figure 6 (right). Surprisingly, these did not result in the same losses of DNA binding activity that were apparent for the full-length protein. Instead, there was a dramatic change in the mobility of the resulting DNA–protein complexes. Thus, protein 181–292 (b) bound to DNA as a ladder of complexes of increasing mobility (lane 5), whereas protein 181–284 (c) showed a single DNA–protein complex with very high mobility (lane 6).

These changes in complex mobility are disproportionate to the differences in protein size and could result from changes in the protein–DNA stoichiometry of each complex or from changes in protein or DNA conformation. The mobilities did not depend on DNA binding, as the positions of the ^{35}S-labeled proteins in native polyacrylamide gels were not substantially changed by the presence or absence of DNA (results not shown). Consequently, we believe that they are a property inherent in these truncated proteins and that they could represent differences in conformation or in the subunit composition of the complexes. The heterogeneity of the 181–292 protein was reproducible in many independent translations and was stable to repeated freeze–thaw cycles. Therefore, the ladder of DNA–protein complexes is unlikely to result from protein degradation but appears to represent alternative conformations or stoichiometries of this protein.

The truncated proteins 181–292 and 181–284 were examined for heteromeric complex formation following cotranslation with full-length USF (Fig. 6, right). As the level of 181–292 (b) protein was increased in this analysis, there was a concurrent increase in the intensity of a broad, intermediate-sized band, which migrated just below a background band that was present in all translations of the full-length protein (lanes 1–5). This novel band apparently reflects the presence of a heteromer, and its diffuse appearance may reflect the heterogeneity of the 181–292 protein. There were no apparent heteromers formed between the full-length USF and the 181–284 protein, which was truncated in the middle of the leucine repeat domain (lanes 6–10). It is possible that weak interactions between the full-length USF and the 181–284 (c) protein are obscured by the background band. In this experiment we added an excess of the 181–284 protein specifically to test this point but could not detect any complexes above background. Furthermore, the 181–284 protein did not form detectable heteromers when cotranslated with the 181–310 protein (results not shown). We interpret these data to mean that the leucine repeat region is required for certain protein–protein interactions but not for DNA binding per se. Because the 181–284 protein apparently cannot form
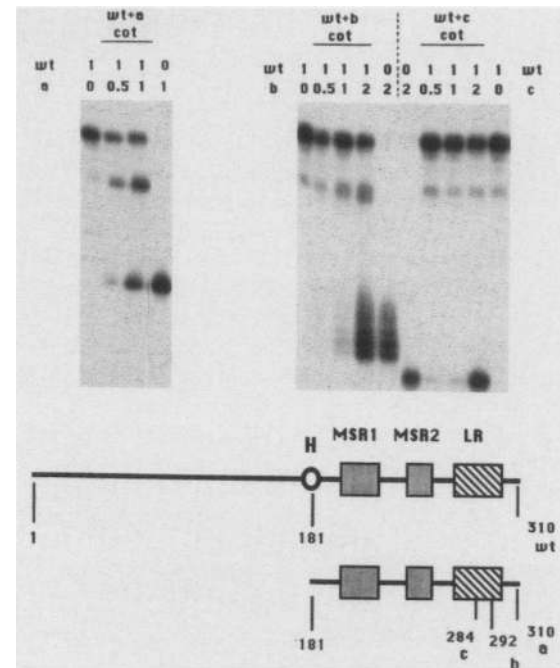

Figure 6. The leucine repeat is required for protein–protein interaction but not DNA binding. Mobility-shift assays, using the ML oligonucleotide as probe. The input (in microliters) of RNA added to each in vitro cotranslation reaction is noted above the lanes. The proteins were wt (1–310), a (181–310), b (181–292), and c (181–284). The free probe is not shown. The map (*bottom*) indicates the sites of template truncation and the location of the hinge (H), MSR1, MSR2, and leucine repeat (LR) domains (see text).

heteromers, and given its relatively rapid mobility in native gels, it may represent a monomeric form of USF. Alternatively, it could represent a dimer whose conformation restricts its ability to form functional heteromers with other proteins; more specifically, it might dimerize through the helix–loop–helix region but only with a protein of equivalent overall conformation.

We compared the DNA sequence specificity of proteins 181–310 and 181–284. If the 181–284 protein binds DNA as a monomer, it may recognize one-half of the symmetrical binding site GGCCACGTGACC, one DNA strand, or one face of the helix. To test these possibilities, unlabeled competitor oligonucleotides were added to the gel retardation assays. These included each single strand that contributes to the ML oligonucleotide used as a probe and a double-stranded oligonucleotide (AAGCTTCGGTGACCGGGATCC) containing half of the binding site (Fig. 7). No differences in affinity for the ML oligonucleotide were seen between proteins 181–310 and 181–284. Furthermore, neither protein was competed by the half-site oligonucleotide (Fig. 7) or by either single strand (results not shown).

Discussion

We described the isolation and preliminary characterization of cDNAs encoding USF, a cellular transcription

factor, and a sequence-specific DNA-binding protein implicated in the regulation of a variety of cellular and viral promoters. Here, we describe the USF DNA-binding domain and its similarity to those of other c-myc-related proteins. We have shown that USF contains two domains, the helix–loop–helix and leucine repeat, which can independently mediate dimerization in other proteins. Thus, USF could, in principle, interact with other cellular factors via either or both of these domains, thus creating unique transcriptional modulators. Functional interactions with basic transcription factors, especially TFIID, and other c-myc-related proteins are currently under study. The interaction between TFIID and USF described previously is of special interest in view of sequence similarities between TFIID and the helix–loop–helix domains of the myc-related proteins (Gasch et al. 1990; Hoffmann et al. 1990).

Heterogeneity of USF protein and RNA

We have shown previously that cellular USF preparations consist of two polypeptides (43 and 44 kD) that have identical DNA-binding activities and co-chromato-

Figure 7. A 102-amino-acid fragment of USF shows the same binding specificity as the full-length protein. Mobility shift assays, using the ML oligonucleotide as probe. Proteins contained amino acids 181–310 (lanes 1–7) or 181–284 (lanes 8–14). No competitor was added to lanes 1 and 8. The amount (in nanograms) and identity of oligonucleotide competitors are indicated above the lanes. The relevant portion of the oligonucleotide sequences is shown at bottom; the full sequence is given in Materials and methods. The symmetrical portion of the ML oligonucleotide is indicated with arrows, and the portion of the ML half-site (ML/2) oligonucleotide that is identical to the ML oligonucleotide probe is underlined.

graph over a variety of conventional resins (Sawadogo 1988; Sawadogo et al. 1988). Each form, when isolated and renatured, gives rise to a unique DNA–protein complex, which is also seen in crude nuclear extracts (Sawadogo et al. 1988). Thus, the coisolation of these forms appears to reflect a similarity in physical properties and not an association between the two forms.

Our inability to sequence the 44-kD form of USF suggested that 43- and 44-kD proteins differ at their amino termini. This is supported by the restricted reactivity of the rabbit antipeptide serum, which recognizes the 43- but not the 44-kD species of cellular USF (M. Sawadogo, unpubl.). These proteins may be the products of different genes, or of differently spliced RNAs arising from a single gene. Alternatively, the differences in size and immunoreactivity might reflect post-translational modifications of a common polypeptide. Isolation of additional related cDNAs and detailed immunological comparisons of USF proteins should resolve this question.

The transcriptional activity that defines the USF protein resides, at least partially, in the protein encoded by our full-length cDNA. Specific antibodies raised against the cloned 43-kD protein both depleted HeLa nuclear extract of UAE-binding activity and specifically reduced transcription from the AdML promoter. In addition, the cloned protein stimulated transcription in vitro. However, studies with both purified and cloned USF components suggested that an additional factor(s) is required to achieve maximal stimulation via the UAE motif on the AdML promoter (Sawadogo 1988; M. Meisterernst, P. Pognonec, P.D. Gregor, and R.G. Roeder, unpubl.). This additional factor(s) is apparently not the 44-kD form of USF, as this protein is clearly present in highly purified preparations of USF that show reduced transcriptional stimulation relative to crude HeLa USF (Sawadogo 1988).

Three classes of cDNAs specific for the 43-kD form of USF, but differing in internal domains, were identified in a human B-cell library (Fig. 1). In our limited survey of the library, the z6b and z18 classes appeared most frequently. The first internal optional domain (OD1) is protein-encoding but is flanked by weak consensus RNA donor and acceptor splice sites. Excision of OD1 in z32 may therefore result from occasional use of these pseudo-splice sites to remove a portion of a larger exon. OD1 may also be encoded by a separate exon, which can be removed by differential RNA splicing. In either case, z32 is expected to encode a nonfunctional protein. The inclusion of OD2 in z18 changes the carboxy-terminal residues and prevents DNA binding. This situation is reminiscent of the Drosophila sex determination genes Sxl and tra (for review, see Baker 1989), in which inclusion of an exon in male flies results in premature translation termination and production of a nonfunctional protein.

Alternative splicing may represent a point of regulation of USF levels, and inclusion of OD2 may dominate in some tissues. Alternatively, the truncated proteins encoded by z18 and z32 may have some distinct biological function. If, for example, the amino-terminal por-
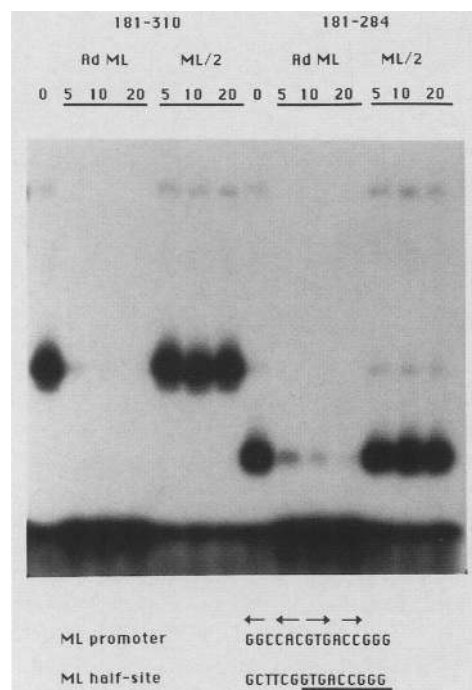
tion of USF contains a transcriptional activation domain, it may be expressed in the shorter proteins and thus modulate the effect of the full-length protein in vivo. Experiments are under way to express each of these cDNAs in vivo and test this hypothesis.

DNA specificity of c-myc-related proteins

The amino acid sequence of its DNA-binding domain identifies USF as a member of the c-*myc*-related family of regulatory proteins. Nearly 20 such proteins have been reported, and all have been implicated in transcriptional activation, differentiation, and/or oncogenesis. Interestingly, some family members have recently been shown to form functional heterodimers via interactions between their helix–loop–helix domains (Murre et al. 1989b; Davis et al. 1990). Specifically, the E12/MyoD heterodimer is thought to be responsible for initiating myogenesis (Davis et al. 1990). Myogenesis is repressed by expression of the c-*myc*-related protein Id, which forms complexes with both *MyoD* and *E12* and prevents their binding to DNA (Benezra et al. 1990).

Interestingly, several c-*myc*-related proteins specifically bind to one of two functionally important and closely related DNA sequences (kE2 and uE3) found within immunoglobulin light (kE2)- and heavy-chain (uE3) enhancers. USF and TFE3 each bind to the uE3 sequence (TGCCACATGACC) and to the ML promoter (GGCCACGTGACC), but not to the kE2 site (GGCCACCTGCCT) (Beckmann et al. 1990; data not shown). In contrast, E12 and E47 bind to the kE2 site and to a closely related sequence in the muscle-specific creatine kinase promoter but not to uE3 (Murre et al. 1989a). Heterodimers E47S/MyoD, E47S/AS-C T3, and da/AS-C T3 are also specific for kE2 (Murre et al. 1989b). These interactions may allow novel couplings of DNA recognition elements with transcriptional activation domains.

We examined the protein sequences for the basis of their related DNA specificities. In the helix–loop–helix domain, similarities between family members are restricted to the conserved residues (shaded in Fig. 3) that are believed to confer the overall structure of the DNA-binding domain and provide no clues for DNA specificity. We therefore focused on the cluster of basic residues adjacent to the first amphipathic helix (amino acids 200–212). In the c-Fos family of proteins, a region rich in basic residues directly precedes a leucine repeat and is essential for DNA binding of c-Fos–c-Jun protein complexes (Kouzarides and Ziff 1988, 1989; Gentz et al. 1989; Turner and Tjian 1989). A structural similarity between the basic regions of the two families has been noted (Prendergast and Ziff 1989). A recent mutational analysis of MyoD showed that the basic region is required for DNA binding and muscle-specific gene activation but not protein dimerization, which required only the helix–loop–helix motif (Davis et al. 1990). The basic regions were therefore compared. Showing a similar overall conservation, USF and TFE3 are identical at 7 of 13 residues and USF and E12 show identity at 6 of 13

residues. However, one block of 6 residues that may correlate with DNA specificity is (V/I)ERRRR in USF and TFE3 or RER(L/V/I)R(V/I) in E12, E47, and da. The corresponding sequence in MyoD, which binds kE2 weakly, is RERRRL. Interestingly, the yeast centromere-binding protein CBF1 both matches USF in this region (VERRRR) and binds to a motif [PuTCACPuTG (Pu = purine)] that encompasses the AdML and uE3 sequence motifs but differs significantly from the kE2 motif (Cai and Davis 1990). The significance of this difference in sequence awaits site-directed mutagenesis of these residues.

In addition to the differences in DNA specificity seen between c-*myc*-related proteins, functional differences exist between proteins with an apparently identical specificity. For example, TFE3 is a *myc*-related protein isolated by virtue of its binding to uE3 (Beckmann et al. 1990). In transient transfection assays, TFE3 stimulates transcription from a reporter gene containing multiple copies of the uE3 motif alone or with other enhancer motifs inserted upstream of a minimal promoter (Beckmann et al. 1990). However, despite its high affinity for the USF-binding site on the AdML promoter, TFE3 does not stimulate transcription from this promoter. A comparison of the activity of each protein on both promoter- and enhancer-dependent constructions should help determine the functional differences between them.

USF has a second dimerization domain

Following the helix–loop–helix domain is a region with periodic leucine residues that has been found in or adjacent to the DNA-binding domains of the Fos–Jun and c-*myc*-related families of proteins (Landschulz et al. 1988a,b). The leucine repeat is required for dimerization and, thus, DNA binding, of C/EBP (Landschulz et al. 1988a, 1989) homodimers and, as predicted (Landschulz et al. 1988a), also allows formation of functional heterodimers (Kouzarides and Ziff 1988, 1989; Sassone-Corsi et al. 1988; Gentz et al. 1989; Schuermann et al. 1989; Turner and Tjian 1989). c-Fos and c-Jun associate in vivo, and both are required for full activation of an AP1-responsive promoter (Chiu et al. 1988; Sassone-Corsi et al. 1988).

In the c-*myc*-related family, the function of the leucine repeat is less clear. The leucine repeat is required for *ras* cotransformation activity and protein tetramerization in c-*myc* (Dang et al. 1989). In contrast, some helix–loop–helix proteins that form heterodimers, such as MyoD/E47S and MyoD/da, lack leucine repeats (Murre et al. 1989b); of these proteins, only da contains a leucine repeat. In addition, we notice that USF and the *myc* proteins are unique among helix–loop–helix proteins in the extreme carboxy-terminal location of their helix–loop–helix and leucine repeat domains. This may suggest a common function for this motif in these two proteins.

We have shown that the leucine repeat in USF is required for protein–protein interactions but not DNA binding per se. In the case of the full-length 310-amino-

acid protein, these protein interactions may be required to expose or configure the DNA-binding domain, because carboxy-terminal truncations that impinge on the leucine repeat disrupt DNA binding significantly. By constrast, deletion of the amino-terminal 180 residues generates a protein that binds DNA independently of the leucine repeat, apparently reflecting constitutive accessibility of the DNA-binding domain. The properties of the protein containing amino acids 181–284, notably its rapid mobility and its inability to interact with longer USF proteins, are most simply interpreted as resulting from changes in conformation and the loss of the leucine repeat. It is also possible that the 181–284 protein binds as a monomer that recognizes the full binding site. This question may be resolved by determining the protein–DNA stoichiometry of the complexes.

Altogether, our data suggest that the leucine repeat is required for USF to attain the conformations needed for both DNA-binding and specific protein–protein interactions. We also infer from these studies that the amino-terminal 180 amino acids of USF impose a conformational restraint on the DNA-binding domain, which must be counteracted by the leucine repeat. We therefore speculate that in vivo, USF function may be repressed or enhanced by specific interactions of other proteins via the leucine repeat or helix–loop–helix domains. In light of this, we attempted to demonstrate specific interactions between USF and several other proteins containing leucine repeat (c-Jun) or helix–loop–helix (MyoD, myogenin, and c-myc) motifs, but no such interactions have yet been seen following cotranslation in vitro (P.D. Gregor, P. Pognonec, M. Sawadogo, and R.G. Roeder, unpubl.).

From kinetic measurements, it was concluded that USF can transfer directly between unlinked DNA-binding sites (Sawadogo 1988). This implies the ability to interact simultaneously with two DNA sequences. We have shown that USF has both a helix–loop–helix and leucine repeat, which each serve as dimerization domains in other proteins. Using these two motifs, the DNA-binding domain of USF might multimerize in some circumstances and bind two DNA target sites by passing through a transient doubly bound intermediate state. The intermediate could, in turn, be stabilized by additional proteins, forming a USF "bridge" between two distant sequences and mediating direct communication between promoter and enhancer regions. In support of this model, the mouse C2 protein, which shares both physical properties and DNA-binding specificity with USF, has been described as a mixture of dimers and tetramers (Peterson and Calame 1989).

Materials and methods

USF protein and oligonucleotide sequences

Five micrograms (100 pmoles, assuming a molecular mass of 43 kD) of USF (mono Q fraction; Sawadogo et al. 1988) were sequenced on an Applied Biosystems model 470 A gas-phase sequenator equipped with a model 120 A phenylthiohydantoin analyzer. The initial sequencing yield was 33 pmoles.

The amino-terminal sequence, with uncertain residues in parentheses, was NH$_2$-(M)KGQQKTAETEEGT(V)QI(Q)(E)G(A/V)(V)(A)(T)(G). The 42-mer best guess oligonucleotide had the sequence 5'-ATG AAG GGC CAG CAG AAG ACA GCT GAG ACA GAG GAG GGC ACC-3'. The sequences of other oligonucleotides described in the text were

Nco: 5'-GTGCCCCCTCACAGAGCCATGGTGAAGGGGCAGCAGAAAACAGC-3'

LR: 5'-TCTTCCACCTGTTGTCG-3'

ML: 5'-CCGGTCGACGGTGTAGGCCACGTGACCGGGTGTAAGCTT-3'
    3'-AGCTGCCACATCCGGTGCACTGGCCCACATTCGAAGGCC-5'

ML/2: 5'-GGCCAAGCTTCGGTGACCGGGATCC-3'
      3'-GGTTCGAAGCCACTGGCCCTAGGGG-5'

DNA sequencing

Double-stranded templates were denatured in alkali, recovered by ethanol precipitation, and sequenced by using the dideoxy chain-termination method (Sanger et al. 1977). Sequenase enzyme and reagents were purchased from American Biochemicals and used according to the manufacturer's instructions. Primers were annealed at 37°C for 20 min, and termination reactions were commonly run at 45°C.

PCR reactions and analysis of products

Reagents were purchased from Perkin Elmer Cetus and used according to the manufacturer's instructions. The oligonucleotide primers, Nco and LR, hybridize to regions 5' of OD1 and 3' of OD2, respectively (Fig. 1). Twenty-five pmoles of each primer and 4.5 μl of a primary phage isolate in SM [10 mM MgCl$_2$, 150 mM NaCl, 20 mM Tris (pH 7.5)] were heated at 95°C for 5 min to burst the phage particles. After cooling, 10× reaction buffer, nucleotides, and Taq polymerase (2.5 units) were added in a final volume of 50 μl. The reaction mixtures were overlaid with mineral oil and subjected to 25 cycles of 1 min at 95°C, 2 min at 55°C, and 3 min at 72°C in a Perkin Elmer Cetus thermocycler. Reactions were cooled to 4°C, and mineral oil was removed by extraction with CHCl$_3$. PCR products were analyzed by agarose gel electrophoresis.

Amino-terminal deletion of amino acids 1–181

A SpeI–AvrII fragment (from polylinker of pBKS$^-$ to nucleotide 657; Fig. 3) was deleted from dI2. The truncated insert was then cloned as an XbaI–EcoRI fragment into the StuI–EcoRI site of the expression vector pT7 β Stu (a generous gift of Richard Treisman), in which the human β-globin gene from pSP64-H β 6 (Krainer et al. 1984) was cloned between the HindIII and PstI sites of pGEM2 and a StuI site was inserted immediately following the ATG of β-globin. The β-globin–USF fusion protein thus contains the first 2 amino acids from β-globin, 3 amino acids from the pBSK polylinker, and amino acids 181–310 from dI2.

In vitro transcription and translation

One microgram of purified, linearized template was transcribed in a 25-μl reaction volume of T7 RNA polymerase (Promega). The RNA was purified by phenol/cholorform extraction and ethanol precipitation, resuspended in TE, and stored at −20°C. The quality and quantity of RNA was monitored by ethidium bromide staining of agarose gels. Additional TE was added if needed to ensure that the RNA concentrations of each sample were approximately equal. From one-twentieth to one-tenth of each transcription reaction was added to a 25-μl in vitro trans-

lation reaction (Promega) containing 2.5 µl of [³⁵S]methionine (Amersham). Proteins were examined by SDS-PAGE prior to their use in other assays.

Mobility-shift assay

Oligonucleotide probes were labeled with [³²P]dCTP and the Klenow subunit of DNA polymerase (New England Biolabs). Mobility-shift reactions (Fried and Crothers 1981) contained 60 mM KCl, 5 mM Tris (pH 7.9), 1 µg poly[d(I-C)], 5% (vol/vol) glycerol, 3 mM DTT, 0.3 ng probe, and 0.5–1.0 µl rabbit reticulocyte lysate in a 10-µl volume. After a 20-min incubation at 30°C, 4 µl of 10% Ficoll was added and samples were loaded onto a [4% (60 : 1 acrylamide/*bis*acrylamide) 0.25× TBE] gel. After electrophoresis at 20 mA, gels were dried and exposed to X-ray film. Two sheets of paper were used to shield the film from the [³⁵S]methionine in the rabbit lysates.

Acknowledgments

References

Baker, B.S. 1989. Sex in flies: the splice of life. *Nature* **340:** 521–524.

Battey, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder. 1983. The human c-myc oncogene: Structural consequences of translocation into the IgH locus in Burkitt lymphomas. *Cell* **34:** 779–787.

Beckmann, H., L.-K. Su, and T. Kadesch. 1990. TFE3: A helix–loop–helix protein that activates transcription through the immunoglobulin enhancer uE3 motif. *Genes Dev.* **4:** 167–179.

Benezra, R., R.L. Davis, D. Lockshon, D.L. Turner, and H. Weintraub. 1990. The protein Id: A negative regulator of helix–loop–helix DNA binding proteins. *Cell* **61:** 49–59.

Bohmann, D., T.J. Bos, A. Admon, T. Nishimura, P.K. Vogt, and R. Tjian. 1987. Human proto-oncogene c-*jun* encodes a DNA binding protein with structural and functional properties of transcriptional factor AP-1. *Science* **238:** 1386–1392.

Brunet, L.J., L.E. Babiss, C.S.H. Young, and D.R. Mills. 1987. Mutations in the adenovirus major late promoter: Effects on viability and transcription during infection. *Mol. Cell. Biol.* **7:** 1091–1100.

Cai, M. and R.W. Davis. 1990. Yeast centromere binding protein CBF1, of the helix–loop–helix protein family, is required for chromosome stability and methionine prototrophy. *Cell* **61:** 437–446.

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.

———. 1987. The major late transcription factor binds to and activates the mouse metallothionein I promoter. *Genes Dev.* **1:** 973–980.

Caudy, M., H. Vaessin, M. Brand, R. Tuma, L.Y. Jan, and Y.N. Jan. 1988. *daughterless*, a *Drosophila* gene essential for both neurogenesis and sex determination, has sequence similarities to *myc* and the *achaete-scute* complex. *Cell* **55:** 1061–1067.

Chiu, R., W.J. Boyle, J. Meek, T. Smeal, T. Hunter, and M. Karin. 1988. The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. *Cell* **54:** 541–552.

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.

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.

Concino, M.F., R.F. Lee, J.P. Merryweather, and R. Weinmann. 1984. The adenovirus major late promoter TATA box and initiation site are both necessary for transcription *in vitro*. *Nucleic Acids Res.* **12:** 7423–7433.

Dang, C.V., M. McGuire, M. Buckmire, and W.M.F. Lee. 1989. Involvement of the "leucine zipper" region in the oligomerization and transforming activity of human c-myc protein. *Nature* **337:** 664–666.

Davis, R.L., H. Weintraub, and A.B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51:** 987–1000.

Davis, R.L., P.-F. Cheng, A.B. Lassar, and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60:** 733–746.

Fried, M. and D.M. Crothers. 1981. Equilibrium and kinetics of Lac repressor interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9:** 6505–6525.

Gasch, A., A. Hoffmann, M. Horikoshi, R.G. Roeder, and N.-H. Chua. 1990. *Arabidopsis thaliana* contains two genes for TFIIID. *Nature* **346:** 390–394.

Gentz, R., F.J. Rauscher III, C. Abate, and T. Curran. 1989. Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. *Science* **243:** 1695–1699.

Gerster, T., C.G. Balmaceda, and R.G. Roeder. 1990. The cell type-specific octamer transcription factor OTF-2 has two domains required for the activation of transcription. *EMBO J.* **9:** 1635–1643.

Hen, R., P. Sassone-Corsi, J. Corden, M.P. Gaub, and P. Chambon. 1982. Sequences upstream from the T-A-T-A box are required *in vivo* and *in vitro* for efficient transcription from the adenovirus serotype 2 major late promoter. *Proc. Natl. Acad. Sci.* **79:** 7132–7136.

Hoffmann, A., E. Sinn, T. Yamamoto, J. Wang, A. Roy, M. Horikoshi, and R.G. Roeder. 1990. Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIIID). *Nature* **346:** 387–390.

Hope, I.A., S. Mahadevan, and K. Struhl. 1988. Structural and functional characterization of the short transcriptional activation region of yeast GCN4 protein. *Nature* **333:** 635–640.

Hough, P.V.C., I.A. Mastrangelo, J.S. Wall, J.F. Hainfeld, M. Sawadogo, and R.G. Roeder. 1987. The gene-specific initiation factor USF (upstream stimulatory factor) bound at the adeno-

virus type 2 major late promoter: Mass and three-dimensional structure. *Proc. Natl. Acad. Sci.* **84:** 4826–4830.

Hu, S.-L. and J.L. Manley. 1981. DNA sequences required for initiation of transcription *in vitro* from the major late promoter of adenovirus 2. *Proc. Natl. Acad. Sci.* **78:** 820–824.

Kouzarides, T. and E. Ziff. 1988. The role of the leucine zipper in the fos-jun interaction. *Nature* **336:** 646–651.

———. 1989. Leucine zippers of *fos, jun* and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature* **340:** 568–571.

Krainer, A.R., T. Maniatis, B. Ruskin, and M.R. Green. 1984. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* **36:** 993–1005.

Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988a. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240:** 1759–1764.

———. 1989. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243:** 1681–1688.

Landschulz, W.H., P.F. Johnson, E.Y. Adashi, B.J. Graves, and S.L. McKnight. 1988b. Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev.* **2:** 786–800.

Lathe, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data. Theoretical and practical considerations. *J. Mol. Biol.* **183:** 1–12.

Lee, R.F., M.F. Concino, and R. Weinmann. 1988. Genetic profile of the transcriptional signals from the adenovirus major late promoter. *Virology* **165:** 51–56.

Lennard, A.C. and J.M. Egly. 1987. The bidirectional upstream element of the adenovirus-2 major late promoter binds a single monomeric molecule of the upstream factor. *EMBO J.* **6:** 3027–3034.

Logan, J. and T. Shenk. 1986. In vivo identification of sequence elements required for normal function of the adenovirus major late transcriptional control region. *Nucleic Acids Res.* **14:** 6327–6335.

Ma, J. and J. Ptashne. 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48:** 847–853.

Mellentin, J.D., S.D. Smith, and M.L. Cleary. 1989. *lyl*-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix–loop–helix DNA binding motif. *Cell* **58:** 77–83.

Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245:** 371–378.

Miyamoto, N.G., V. Moncollin, M. Wintzerith, R. Hen, J.M. Egly, and P. Chambon. 1984. Stimulation of *in vitro* transcription by the upstream element of the adenovirus-2 major late promoter involves a specific factor. *Nucleic Acids Res.* **12:** 8779–8799.

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.

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.

Muller-Immergluck, M.M., W. Schaffner, and P. Matthias. 1990. Transcription factor Oct-2A contains functionally redundant activating domains and works selectively from a promoter but not from a remote enhancer position in nonlymphoid (HeLa) cells. *EMBO J.* **9:** 1625–1632.

Murre, C., P. Schonleber-McCaw, and D. Baltimore. 1989a. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless, MyoD,* and *myc* proteins. *Cell* **56:** 777–783.

Murre, C., P. Schonleber-McCaw, H. Vaessin, M. Caudy, L.Y. Jan, C.V. Cabrera, J.N. Buskin, S.D. Hauschka, A.B. Lassar, H. Weintraub, and D. Baltimore. 1989b. Interactions between heterologous helix–loop–helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58:** 537–544.

Pabo, C.O. and R.T. Sauer. 1984. Protein-DNA recognition. *Annu. Rev. Biochem.* **53:** 293–321.

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.

Peterson, C.L. and K. Calame. 1989. Proteins binding to site C2 (uE3) in the immunoglobulin heavy-chain enhancer exist in multiple oligomeric forms. *Mol. Cell. Biol.* **9:** 776–786.

Prendergast, G.C. and E.B. Ziff. 1989. DNA-binding motif. *Nature* **341:** 392.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Sassone-Corsi, P., L.J. Ransone, W.W. Lamph, and I.M. Verma. 1988. Direct interaction between fos and jun nuclear oncoproteins: Role of the "leucine zipper" domain. *Nature* **336:** 692–695.

Sawadogo, M. 1988. Multiple forms of the human gene-specific upstream stimulatory transcription factor. II. DNA binding properties and transcriptional activity of the purified HeLa upstream stimulatory factor. *J. Biol. Chem.* **263:** 11994–12001.

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.

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.

Schuermann, M., M. Neuberg, J.B. Hunter, T. Jenuwein, R.-P. Ryseck, R. Bravo, and R. Muller. 1989. The leucine repeat motif in fos protein mediates complex formation with jun/AP-1 and is required for transformation. *Cell* **56:** 507–516.

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.

Smale, S.T. and D. Baltimore. 1989. The "initiator" as a transcription control element. *Cell* **57:** 103–113.

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.

Turner, R. and R. Tjian. 1989. Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. *Science* **243:** 1689–1694.

Villares, R. and C.V. Cabrera. 1987. The *achaete-scute* gene complex of D. melanogaster: Conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* **50:** 415–424.

Workman, J.L., R.G. Roeder, and R.E. Kingston. 1990. An upstream factor, USF (MLTF), facilitates the formation of preinitiation complexes during in vitro chromatin assembly. *EMBO J.* **9:** 1299–1308.

Wright, W.E., D.A. Sassoon, and V.K. Lin. 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56:** 607–617.

Yu, Y.-T. and J.L. Manley. 1984. Generation and functional analyses of base-substitution mutants of the adenovirus 2 major late promoter. *Nucleic Acids Res.* **12:** 9309–9321.

References	This article cites 62 articles, 21 of which can be accessed free at: **http://genesdev.cshlp.org/content/4/10/1730.full.html#ref-list-1**
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**