A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif

Robert Duncan,1 Leonard Bazar,2,3 Greg Michelotti,1,3 Takeshi Tomonaga,1 Henry Krutzsch,1 Mark Avigan,2 and David Levens1

1Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892 USA; 2Departments of Pathology and Medicine, Georgetown University, Washington, D.C. 20007 USA

The far upstream element (FUSE) of the human c-myc proto-oncogene stimulates expression in undifferentiated cells. A FUSE-binding protein (FBP) is present in undifferentiated but not differentiated cells. Peptide sequences from the purified protein allowed cloning of cDNAs encoding FBP. Expression of FBP mRNA declined upon differentiation, suggesting transcriptional regulation of FBP. Features in the FBP cDNA suggest that FBP is also regulated by RNA processing, translation, and post-translational mechanisms. Both cellular and recombinant FBP form sequence-specific complexes with a single strand of FUSE. Transfection of FBP into human leukemia cells stimulated c-myc-promoter-driven expression from a reporter plasmid in a FUSE-dependent manner. Deletion and insertion mutagenesis of FBP defined a novel single-strand DNA-binding domain. Analysis of the primary and predicted secondary structure of the amino acid sequence reveals four copies of a reiterated unit comprised of a 30-residue direct repeat and an amphipathic α-helix separated by an 18- to 21-residue spacer. The third and fourth copies of this repeat-helix unit constitute the minimum single-stranded DNA-binding domain. To determine whether the FUSE site, in vivo, possesses single-strand conformation, and therefore could be bound by FBP, cells were treated with potassium permanganate (KMnO4) to modify unpaired bases. Modification of genomic DNA in vivo revealed hyperreactivity associated with single-stranded DNA in the FUSE sequence and protection on the strand that binds FBP in vitro. The role of single-stranded DNA and single-strand binding proteins in c-myc regulation is discussed.

[Key Words: cDNA clone; myc expression; DNA-binding protein; differentiation; potassium permanganate (KMnO4)]

Received August 11, 1993; revised version accepted January 5, 1994.

The c-myc proto-oncogene plays a central role in normal cell proliferation [Spencer and Groudine 1991] and programmed cell death [Shi et al. 1992], and its deregulation contributes to the formation of a variety of tumors [Bishop 1983; Cole 1986; Cory 1986]. Even small changes in c-myc abundance can significantly alter cell growth [Shichiri et al. 1993]. A multitude of signals can alter c-myc expression, thereby affecting a cell’s decision to proliferate; many of these signals may exert their influence through cis-acting elements within or near the c-myc gene, some of which have been shown to bind nuclear proteins [Marcu et al. 1992]. Several studies have demonstrated that the chromatin surrounding the c-myc gene is sensitive to cleavage with nucleases at discrete sites [Siebenlist et al. 1984; Grosso and Pitot 1985b; Bentley and Groudine 1986a], presumably reflecting perturbation of the DNA by these bound proteins. The presence, absence, or intensity of some nuclease hypersensitive sites has been correlated with the expression of c-myc [Shubach and Groudine 1984; Siebenlist et al. 1984, 1988; Bentley and Groudine 1986a, 1986b]. Understanding how the c-myc regulatory proteins interact with each other and the basal transcription apparatus to produce an integration of multiple signals requires a more complete characterization of the elements involved.

Down-regulation of the c-myc proto-oncogene occurs in the human promonomyelocytic leukemia cell line HL60 and human monoblastic line U937 upon induction of differentiation [Dony et al. 1985; Grosso and Pitot 1985a; Watanabe et al. 1985; Bentley and Groudine 1986a, Eick and Bornkamm 1986]. This suppression of c-myc expression occurs by two mechanisms, within 3

3These authors made an equal contribution to this work.
hr there is a block to elongation that can be reversed by removal of the differentiation agent. Subsequently, transcriptional initiation ceases, coinciding with irreversible commitment to the differentiation pathway (Siebenlist et al. 1988). A far upstream element (FUSE), required for maximal transcription of c-myc, binds a factor that is present in extracts of undifferentiated cells but disappears upon differentiation (Avigan et al. 1990). The disappearance of this binding activity occurs 24 hr after addition of the differentiation agent, coinciding with the loss of initiation of c-myc transcription. Deletion of the FUSE site results in a fivefold reduction in CAT expression from a plasmid with 3.2 kb of c-myc sequence fused to CAT in transfection experiments (Avigan et al. 1990). However, this site differs from other described positive regulatory elements for myc in a number of ways. Despite its position a long distance from the transcription start site (−1500 bp relative to the myc P1 promoter), the FUSE element will not act as a traditional enhancer; multiple copies inserted upstream of a heterologous promoter fail to stimulate expression in transfection experiments. In contrast, when the FUSE site is present with additional c-myc regulatory sequences, specific stimulation of the c-myc promoter is observed, indicating that FUSE works in concert with other myc elements. This interpretation is supported by stably transfected c-myc promoter CAT constructs that fail to down-regulate with differentiation of U937 cells unless both a distal domain, encompassing FUSE, and a proximal region are both present (Skerka et al. 1993). These features suggest that the factor or factors binding to this site may act by an unusual mechanism.

This paper describes the isolation of cDNA clones that encode a FUSE-binding protein (FBP). The recombinant protein contains a repeated structure that was demonstrated to be required for DNA binding and, hence, defines a new DNA-binding motif. Unexpectedly, the recombinant as well as purified cellular FBP bind preferentially and specifically to single-stranded DNA possessing the sequence of the noncoding strand of the FUSE site. FBP stimulates expression in a FUSE-dependent manner, suggesting an important regulatory role for this protein. Evidence for perturbed DNA structure in the region of the FUSE site, in vivo, is presented, and the potential role of single-strand DNA and single-strand binding proteins in transcription is discussed.

Results

Isolation of cDNAs encoding FBP

A 70-kD FBP was purified from undifferentiated HL60 cells by oligonucleotide affinity chromatography (Wu et al. 1987), the protein was eluted from an SDS–acylamide gel, and the amino acid sequence of internal peptides was determined [L. Bazar, R. Duncan, D. Levens, and M. Avigan, in prep.]. The peptide sequences enabled the design of degenerate oligonucleotide PCR primers that were used to amplify a cDNA template prepared from undifferentiated U937 cell total RNA. Two amplified products contained internal open reading frame-encoding segments identical to peptide sequences from purified FBP, confirming that the authentic cDNA had been obtained.

The PCR fragments hybridized to a single 2.6-kb RNA that disappeared after dimethylsulfoxide (DMSO)-induced differentiation [Fig. 1A] consistent with the loss of FUSE-binding activity after DMSO treatment. Expression of the 2.6-kb transcript declined sharply after 24 hr of treatment with TPA and was undetectable at 48 hr (Fig. 1B), coinciding with [1] the disappearance of the FUSE-binding activity, [2] a dramatic decrease in c-myc transcriptional initiation, and [3] an irreversible commitment to terminal differentiation. Hence, FBP mRNA and

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Differentiation regulated expression of FBP mRNA. Polyadenylated RNA (10 μg) from cultured cells was separated on a 1% agarose–formaldehyde gel, transferred to nitrocellulose, and hybridized with a 32P-labeled probe from FBP, β-actin, or human c-myc cDNAs. (A) RNA prepared from HL60 cells grown for 48 hr in the presence (+) or absence (−) of 1.3% DMSO. The blot was hybridized with the FBP probe, and was stripped and rehybridized with actin. Identical samples from treated or untreated cells are alternately loaded for comparison. (B) RNA prepared from HL60 cells cultured as indicated with or without 50 ng/ml of 12-O-tetradecanohlphorbol 13-acetate (TPA). The blot was hybridized with FBP then sequentially stripped and rehybridized with actin and c-myc.
binding activity are shut off synchronously during differ-
entiation.

A full-length FBP cDNA sequence was assembled from overlapping clones obtained from three libraries using the PCR fragments as probes (Fig. 2A). The composite 2385-bp cDNA contains 26 bp of 5' untranslated sequence, 1932 bp of open reading frame, and 384 bp of 3'-untranslated sequence, including a poly(A) addition signal and 60 nucleotides of poly(A) [Fig. 2B]. Restriction fragments throughout the cDNA hybridize to the single 2.6-kb FBP mRNA [data not shown]. Although the 5'-untranslated region of the message appears to be unusually short and contains no in-frame stop codons, the initi-
tiator methionine is contained in a 9 out of 10 match with the Kozak consensus [Kozak 1991]. The deduced polypeptide, 644 amino acids in length, has a calculated molecular mass of 67.5 kD [Fig. 2C] consistent with the molecular mass of the purified protein. All of the remaining sequenced peptides were found in the deduced polypeptide [single underline in Fig. 2C].

The cDNA sequence has features that suggest multiple modes of FBP regulation. There is an unusual codon bias. For every amino acid with a degenerate codon, except glutamine, the FBP sequence avoids the codons preferred in a survey of 2681 human genes [Wada et al. 1992] suggesting translational regulation. The existence of alternate isoforms of FBP generated by regulation of RNA processing is suggested by multiple independent clones, half of which exhibit a precise deletion of the 3 nucle-

otides encoding serine 97 (see Fig. 2B, C). The surrounding sequence is consistent with an alternate splice ac-
ceptor site [Mount 1982] that could generate this dele-
tion.

DNA-binding properties of FBP

Unexpectedly, the purified human FBP bound preferen-
tially to the noncoding strand of the FUSE site as a sin-
gle-stranded oligonucleotide. In an electrophoretic mo-
bility-shift assay [EMSA] with labeled double-stranded FUSE oligomer, labeled coding strand, or labeled noncod-
ing strand as probes, the protein fraction that was eluted from the FUSE oligonucleotide column bound to the noncoding strand only. No significant complex for-
ation was observed with the annealed double-stranded FUSE oligonucleotide, the coding strand, or other single-stranded oligonucleotides as probe [Fig. 3A, lanes 7–12].

To determine whether these DNA-binding properties were exhibited by cloned FBP, the composite cDNA se-
sequence was subcloned into a bacterial expression vector and glutathione S-transferase–FBP fusion protein [GST–

FBP] was purified from bacterial extracts. The recombi-
nant FBP fusion protein shows the same binding prefer-
ance as the native FBP fraction [Fig. 3A, lanes 13–18]. To ex-
amine the specificity of the single-stranded binding of native and recombinant FBP, binding to labeled noncod-
ing strand was challenged with a panel of competitors [Fig. 3B]. As little as a fivefold molar excess noncoding FUSE oligonucleotide competed for binding to either pu-
rified cellular or recombinant FBP [lanes 4, 12]. In con-
trast, greater excesses of the following oligonucleotides showed no significant binding to either protein prepara-
tion: double-stranded FUSE oligonucleotides, a mutant noncoding strand, the coding or noncoding strands of another cis element found in the myc gene [lanes 7–10, 15–18], or the annealed products of these oligonucleotides [data not shown].

FBP stimulates c-myc promoter-mediated expression

The FUSE site is required for maximal activation of the c-myc promoter in U937 cells. Chloramphenicol acetyl-
transferase [CAT] expression from a transiently trans-
sected reporter plasmid containing 3.2 kb of human c-myc regulatory sequence fused to a CAT gene [myc–
CAT] is reduced when 4 nucleotides within the FUSE site (Δ-4 nt) are deleted [Avigan et al. 1990, Fig. 4A]. The same reduction in activation occurs if the reporter plas-
mid has a 68-bp deletion centered on the FUSE site (ΔFUSE–CAT, Fig. 4A, top). The difference between the FUSE-containing and FUSE-deleted plasmids was more pronounced in U937 cells suggesting that other cell lines may not provide cooperating components necessary for FUSE activity [Skerka et al. 1993]. Therefore, U937 cells were used to further investigate the effect of FBP on FUSE-mediated activation. However, this cell line already expresses some FUSE-binding activity [Avigan et al. 1990] and FBP message [R. Duncan, unpubl.]. To ob-
serve an effect of FBP produced by a transfected expres-
sion plasmid, conditions were sought in which endoge-
nous factors were limiting. Transfecting increasing amounts of the myc–CAT reporter plasmid above the optimum at 1 μg leads to a decline in CAT expression and convergence in the amount of expression from the FUSE-containing [myc–CAT] and the mutated plasmid (Δ-4 nt; Fig. 4A, bottom), suggesting that at higher plas-
mid concentrations these limiting conditions have been reached. Consequently, transfections were performed with 10 μg of the myc–CAT reporter plasmids and 10 μg of expression plasmids [Fig. 4B]. At this level, in the ab-

ence of exogenous FBP [Fig. 4B, vector alone lanes], only a small difference in CAT expression was observed be-


tween the FUSE-containing myc–CAT plasmid (myc–
CAT) and the reporter plasmid with a 68-bp deletion centered on the FUSE site (ΔFUSE myc–CAT). In the presence of the FBP expression plasmid [FBP-sense], the myc–CAT plasmid gave a fivefold higher level of CAT activity than in the presence of the vector alone. In con-

trast, this level of stimulation did not occur with the ΔFUSE myc–CAT reporter plasmid, nor with the expres-
sion plasmid containing the FBP cDNA in the reverse oriention [FBP antisense]. The minor increase in the ΔFUSE myc–CAT expression with FBP cotransfection could be the result of uncharacterized cis-acting ele-
ments present in the myc regulatory sequence. Cotrans-
fection of an FBP expression plasmid that has the se-
quence deleted encoding the repeated DNA-binding mo-
tifs [see below] does not alter expression from the myc–
CAT or the FUSE-deleted myc–CAT reporter [data not shown].
Analysis of the primary sequence of FBP

FBP shows no significant homology to proteins with known DNA-binding motifs in a search of the GenBank data bases. However, the primary amino acid sequence of FBP has distinct structural features that could constitute a DNA-binding domain, as well as other features with potential functional significance. A Pustell matrix self-comparison of FBP revealed three domains, amino-ter-

Figure 2. Primary structure of FBP cDNA clones. [A] Maps of FBP cDNA sequences. The open box in the composite cDNA indicates the coding region, and Ap, the poly[A] tail. The positions of PCR products and clones are shown; solid lines represent vertically aligned identical sequence; broken lines represent nonhomologous sequence. Clone names [left] indicate source or library of origin. The nonhomologous sequences in the HL60 clones, BJAB-1, and PBL-3 most likely are from reverse-transcribed, unspliced pre-mRNA or extraneous cDNA fragments ligated together during the preparation of the libraries. The open circle on clone BJAB-2 indicates the position of a 3-bp deletion. [B] Nucleotide sequence of the composite cDNA clone of FBP and predicted amino acid sequence. Nucleotides and amino acids are numbered at the right of each line. [C] Deduced amino acid sequence of FBP. Peptides identified by amino acid sequencing of cyanogen bromide/trypsin-cleaved, purified human FBP are indicated by a single underline. Arrowheads mark the position of a PCR product. Open boxes enclose predicted amphipathic α-helices; solid boxes enclose a 30-residue repeated motif; the serine residue indicated by an asterisk (*) is deleted in a variant form of the FBP message; the double underlined residues comprise an 8-amino-acid direct repeat; the I and Y above positions 149 and 618, respectively, are alternate amino acids encoded by the clone PBL-2. Abbreviations for the amino acid residues are A (Ala), C (Cys), D (Asp), E (Glu), F (Phe), G (Gly), H (His), I (Ile), K (Lys), L (Leu), M (Met), N (Asn), P (Pro), Q (Gln), R (Arg), S (Ser), T (Thr), V (Val), W (Trp), and Y (Tyr).
FBP stimulates myc expression

Figure 2. [See facing page for legend.]
Figure 3. FBP binds to the FUSE DNA sequence. (A) Purified human and recombinant FBP bind to the noncoding strand of FUSE. EMSA was performed using the indicated $^{32}$P-labeled oligonucleotides [Probe]. Probes were incubated with human FBP purified from HL60 cells by oligonucleotide affinity chromatography (Purified FBP) or the polypeptide encoded by the cDNA expressed as a GST fusion protein (Recombinant FBP). DNA–protein complexes were separated from free probe (F) on a nondenaturing polyacrylamide gel and visualized by autoradiography. The DNA strand that carries the coding sequence of the mRNA in the transcribed region is designated the coding strand of the oligonucleotides, and the complement is designated the noncoding strand, despite their upstream location. FUSE double strand was formed by annealing FUSE coding with noncoding strand oligonucleotides. Probe sequences are described in Materials and methods. (B) Recombinant FBP and purified human FBP display similar DNA-binding specificity. EMSA of recombinant FBP and purified FBP was performed using radioactive FUSE noncoding strand as probe. Proteins were incubated with the probe or a combination of the probe and the indicated molar excess of unlabeled FUSE noncoding strand [lanes 4–6, 12–14] or at least a 10-fold molar excess of the other indicated competitors and processed as in A. GST alone did not bind to the probe [lane 2].

minal, central, and carboxy-terminal, each containing internally repeated sequences [Fig. 5A].

The amino-terminal domain is comprised of 106 amino acids featuring a string of 11 repeated glycine residues and the first of 5 predicted amphipathic α-helices found in FBP. Following the helix, residues 63–106 are enriched for glutamine (16%) relative to the whole protein (9%).
FBP stimulates myc expression

Figure 4. The dose-dependent positive effect of the FUSE site is augmented by FBP. [A (Top)] CAT expression from a reporter plasmid bearing 3.2 kb of c-myc regulatory sequence fused to a CAT gene is decreased by deletions in the FUSE site in transient transfections. U937 cells were transfected using a DEAE-dextran procedure with a plasmid containing the native myc sequence (lanes labeled myc-CAT) or a plasmid that differed by a 4-nucleotide deletion in the FUSE site (lanes labeled Δ-4 nt) described previously as pMP CAT and pMP CATΔ1, respectively [Avigan et al. 1990]. The same decrease occurred in transfections with the pMP CATΔFUSE plasmid with the entire FUSE site deleted (ΔFUSE-CAT lanes). The amount of plasmid used is shown above each set. [Bottom] CAT expression declines and the difference between the FUSE-containing and FUSE-deleted constructs diminishes as the amount of plasmid (shown above each pair) is increased from 1 µg to 5 µg. Transfections were performed as above. [B] FBP stimulates expression in a FUSE-dependent manner. Full-length FBP was inserted into an expression vector downstream from the cytomegalovirus enhancer/promoter in both the sense and antisense orientations. These plasmids or the vector alone were transfected into U937 cells with another plasmid containing a CAT gene under the control of myc regulatory sequence (myc-CAT) or the reporter plasmid with the FUSE site deleted (ΔFUSE myc-CAT). U937 cells [5 × 10⁶] were electroporated with 10 µg of each plasmid in 250 µl, transferred to 8 ml of culture medium, and incubated 48 hr before harvesting for CAT assays. Two independent transfections are shown for each plasmid combination. The reporter plasmids in this experiment, pMP Nsi CAT and pMP Nsi CATΔFUSE, were modifications of pMP CAT and pMP CATΔFUSE, respectively [see Materials and methods].

The central and largest FBP domain contains structures sufficient for sequence-specific binding because a recombinant protein comprised of only this domain bound DNA specifically [see below]. The central domain is made up of four evenly spaced units, each unit contains a highly conserved 30-residue segment termed the FBP repeat [Fig. 5B], followed by a potential amphipathic α-helix [Fig. 5C]. Each FBP repeat is comprised of three structures in sequential order: a region of predicted β-sheet separated by a turn at conserved glycines 13 and 14 from a segment of α-helix as predicted by Chou–Fasman and Robson–Garnier analysis [Chou and Fasman 1978; Garnier et al. 1978]. The invariant glycine at position 25 may not disrupt helix formation because of the hydrophilic character of neighboring residues in the predicted helix [Serrano et al. 1992]. This sheet–turn–helix pattern evokes the DNA-binding domain of the papovavirus E2 protein [Hegde et al. 1992]. The FBP repeat shares homology with a repeated sequence in a protein that was isolated from heterogeneous nuclear ribonucleoprotein complexes, hnRNP protein K [Matunis et al. 1992, Fig. 5B]. Recently, hnRNP protein K was shown to bind and trans-activate the “CT” element located between positions –151 and –100 relative to the c-myc P1 promoter [Takimoto et al. 1993]. This related protein displays different DNA-binding specificity than FBP, forming a tight, specific complex with a single strand of the CT element but not with the complementary strand [Takimoto et al. 1993; Fig. 3B]. FBP/hnRNP protein K-like amino acid repeats have been reported in several nucleic acid-associated proteins but have not been functionally characterized [Siomi et al. 1993]. Despite the homology to hnRNP protein K, FBP is unlikely to be an RNA-binding protein. FBP, unlike hnRNP protein K, is not found in hnRNP complexes [M. Matunis and G. Dreyfuss, pers. comm.]. It is localized in the nucleus exclusive of the nucleolus, the site of rRNA synthesis [data not shown]. Furthermore, FBP lacks any well-defined RNA-binding motif, and experiments have not elicited specific RNA binding [G. Michelotti and D. Levens, unpubl.].

An amphipathic helix follows the FBP repeat in each of the four evenly spaced units, separated by a spacer of 18–21 residues. The amphipathic character of the four central domain helices extends uniformly over their 16- to 18-residue lengths with the exception of hydrophilic...
residues at position 2 (Fig. 5C). The lack of a hydrophobic residue at this position deviates from the heptad repeat pattern shown to be important in coiled–coiled interactions (Crick 1952). The helices are not flanked by basic regions as in basic–coiled–coil (Landschulz et al. 1988, Harrison 1991) or helix–loop–helix motifs (Murre et al. 1989). Thus, the repeat-helix unit found in the central domain of FBP does not resemble known DNA-binding motifs.

The carboxy-terminal domain is separated from the central domain by a highly flexible, proline/glycine-rich segment. This domain is also glutamine rich (22/140 = 15.7% from amino acid 505 to the carboxyl terminus) and tyrosine rich (13/140 = 9.3%) including repeated tyrosine–dyad motifs, some of which conform to a tyrosine phosphorylation recognition site (Cooper et al. 1984).

**Identification of the DNA-binding domain**

To define a minimum DNA-binding motif, truncations were constructed as well as short insertions to disrupt
were expressed as bacterial fusion proteins, purified, and characterized by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) [Fig. 6C, bottom]. Full-length, truncated, and mutant constructs were tested for DNA binding with EMSAs. Truncated proteins deleting the amino-terminal region through the first (or first and second) repeating units and the carboxy-terminal region extending into the proline/glycine-rich sequence [Fig. 6A, constructs 145–511, 278–511, and 278–474] bound specifically to the noncoding strand of the FUSE oligonucleotide as did the full-length FBP [Fig. 6C, top]. In contrast, the removal of unit four and the rest of the carboxyl terminus (construct 278–372), or removal of the amino-terminal portion of repeat 3 (construct 298–511), abrogated binding. These truncations suggest that at least two repeating units are required to constitute a DNA-binding domain.

Insertion mutants further demonstrated the importance of repeat-helix units 3 and 4 for DNA binding. Three to six amino acids were introduced at eight positions throughout the 278–511 construct [Fig. 6B], otherwise preserving the reading frame. An insertion at the junction between GST and the FBP sequence (insertion mutant number 1) does not affect binding, nor do insertions into the region between helix 4 and repeat 4 (number 6) or the Pro-Gly region (number 8). In contrast, insertion 2 in repeat 3, insertions 3 and 4 in helix 4, and insertion 7 in repeat 4 all resulted in mutant proteins failing to bind DNA. Insertion number 5, near the end of helix 4 reduced but did not eliminate binding. The multiple complexes separated on the EMSA gel for some of the protein preparations may be attributed to prematurely terminated or bacterially processed FBP fragments retaining DNA-binding capability; all of the polypeptides shown in Figure 6C, bottom (except GST in lane 2), are reactive with affinity-purified anti-FBP by immunoblot analysis [data not shown].

**FUSE DNA displays single-strand properties in vivo**

The binding of FBP to the noncoding strand of FUSE in vitro provoked an examination of the state of FUSE in vivo. Earlier studies demonstrated chromatin sensitive to SI nuclease, which is specific for single-strand nucleic acid, upstream of the human c-myc gene (Grosso and Pitot 1985); however, these sites were not mapped with sufficient accuracy to relate them to FUSE. Potassium permanganate, which reacts preferentially with thymine in single-strand regions, has been used to probe DNA conformation, in vivo and in vitro, with single-base resolution. Permanganate oxidized genomic DNA, following β-elimination with piperidine, can be subjected to ligation mediated PCR and primer extension analysis to map the sites of base modification and strand cleavage. This approach has been used recently to demonstrate the separation of strands in transcription bubbles caused by an elongationally arrested RNA polymerase, in vivo, downstream of promoter P2 of the human c-myc gene (Krumm et al. 1992) and the promoter of Drosophila hsp70 (Giardina et al. 1992).

Whole cells or naked genomic DNA were treated with KMnO₄ to compare the patterns of reactivity of DNA, in vivo or in vitro. Sites of modification were revealed using LM-PCR with primers appropriate to display the FUSE region. To provide unambiguous markers with single-nucleotide resolution for the identification of reactive bases, human c-myc DNA was reacted with hydrazine, to generate a C + T ladder, or cleaved with AvaI, which cuts in the FUSE element, and these products were also subjected to ligation-mediated polymerase chain reaction (LM–PCR) and displayed by primer extension. An example of the application of this approach to the FUSE region is shown in Figure 7, A and B. Figure 7C summarizes the pattern of permanganate reactivity noted in several independent experiments.

In vivo, the coding strand exhibited numerous hyperreactive thymidine residues extending from thymidine 778 in the FUSE [numbering from the HindIII site 2329 bases upstream of P1] to 823. Cytidines 789 and 790 in FUSE were also hyperreactive, in vivo, compared with naked DNA; reactivity of KMnO₄ with C and G residues has been noted in conformationally perturbed DNA [Rubin and Schmid 1980; Gibson et al. 1993]. No protected nucleotides nor other evidence of protein–DNA interactions were seen on this strand. The overall pattern was consistent with an open single strand extending from FUSE toward P1.

In contrast, the noncoding strand, in vivo, was predominantly hyporeactive when compared with naked, genomic DNA. Several nucleotides, in the DNA segment bound by FBP in vitro, appeared to be protected from reaction with permanganate. Thymidines 775, 777, and 804, cytidine 792, and guanidine 790 and 791 all displayed reduced reactivity. Only thymidine 800 was hypersensitive in vivo. The simultaneous presence of hypersensitivity to permanganate on the coding strand and protection of the noncoding strand is consistent with the notion that FBP binds to the noncoding strand of FUSE while displacing its partner in vivo.

**Discussion**

A composite cDNA has been identified encoding FBP, a sequence-specific, single-stranded DNA-binding protein that increases c-myc promoter-driven expression through the FUSE. Identity of the polypeptide encoded by the cDNA with FBP is proven by the presence of the sequences of 10 peptides derived from the oligonucleotide affinity-purified cellular protein. Recombinant FBP and purified human FBP exhibit identical DNA-binding specificity for the noncoding strand of the FUSE site.

**Regulation of FBP**

Several lines of evidence indicate multiple modes of regulation of FBP. The decrease in FBP mRNA expression with differentiation indicates that the disappearance of the FUSE-binding activity may be regulated transcrip-
Figure 6. Mutational analysis of the minimum DNA-binding domain of recombinant FBP. (A) Maps of FBP and truncation mutants containing amino acids inclusive of the corresponding position numbers are shown at the left. Each construct's ability to bind (+) or failure to bind (−) the FUSE DNA sequence is indicated at right. The open boxes labeled GST indicate the glutathione S-transferase fragment present in the fusion proteins. Other symbols as in Fig. 5A. (B) Diagram of FBP mutants containing amino acids 278–511 altered by insertions and deletions (∆) of the amino acids shown at the indicated locations without changing the remainder of the amino acid sequence. The DNA-binding property of each insertion mutant is indicated by a +, −, or R (reduced binding) at right. (C) Characterization of full-length FBP and mutants shown in A and B. (Bottom) A Coomassie-stained SDS–polyacrylamide gel of bacterially expressed, glutathione-agarose-purified GST and GST fusion constructs of FBP. Molecular size markers indicate the position of protein standards included on the gel. (Top) An EMSA of the same protein samples using labeled FUSE noncoding strand probe as described in Fig. 3. Variable migration of the complexes on the nondenaturing EMSA gel may result from truncation or disrupted folding of the mutant proteins.
Phosphorylated protein kinase C recognition motifs in the predicted amino acid sequence encoded by FBP defines a novel DNA-binding motif. The presence of tyrosine kinase, as well as particular tRNA pools, are correlated in bacteria and have been shown to be sensitive to growth conditions (Emilsén et al. 1992; Gupta et al. 1993). A similar phenomenon could be predicted for eukaryotic cells. Finally, the abundance of active FBP may be determined by translation efficiency. The highly unusual codon bias in FBP suggests that its translation might be dependent on the abundance of minor tRNA species. Preferred codon use and abundance of particular tRNA pools are correlated in bacteria and have been shown to be sensitive to growth conditions (Emilsén et al. 1993). A similar phenomenon could be predicted for eukaryotic cells. Finally, the abundance of active FBP may be regulated by post-translational modification. The presence of tyrosine kinase, as well as protein kinase C recognition motifs in the predicted amino acid sequence, suggests that cellular FBP may be phosphorylated. Protein phosphorylation is one mechanism of modifying transcription factor activity (Smeal et al. 1992; Gupta et al. 1993).

**FBP defines a novel DNA-binding motif**

Analysis of the primary amino acid sequence encoded by the FBP cDNA reveals four repeated units, each comprised of a 30-residue conserved sequence, and an amphipathic α-helix separated by an 18- to 21-residue spacer. The regions between the units are rich in prolines and glycines that may provide flexibility to allow each repeated segment to fold independently. Deletion and insertion mutagenesis of the FBP polypeptide shows that repeated units 3 and 4 constitute a minimum DNA-binding domain, still displaying specificity toward FUSE. Disruption of repeated segments predicted to form either β-sheet or amphipathic helix abrogate DNA binding, indicating that correct folding of an extensive region is likely to be necessary for FBP to interact with DNA. Amphipathic α-helices are important for dimerization in other DNA-binding proteins such as myc (Murre et al. 1989), C/EBP (Landschulz et al. 1988), and EBF (Hagman et al. 1993). Attempts to demonstrate FBP dimerization have been unsuccessful, suggesting that the FBP amphipathic helices promote the intramolecular folding required to form a functional DNA-binding domain. Some of the structural features of FBP that are important for DNA binding are shared with hnRNP protein K; in particular, the conserved sequence is similar to a motif repeated in hnRNP protein K three times; however, in hnRNP protein K the repeated elements are not followed by amphipathic helices. Because hnRNP protein K and FBP interact with dissimilar sequences in single-stranded DNA, the conserved residues most likely pro-

![Image of gel electrophoresis](https://example.com/image.png)
vide only an architectural framework and specificity is conferred by other features. Additional studies will demonstrate whether the shared structures function equivalently in the two proteins.

The presence of four sets of the repeat-helix unit in FBP, when only two are required for FUSE binding, suggests that this protein has the potential to form at least two binding sites. If present, a dual-binding capability could allow FBP to facilitate DNA-loop formation juxtaposing distant elements, a mechanism consistent with the far upstream location and the requirement for additional regulatory sequence to observe the stimulatory effect of the FUSE site. In a similar manner, by bending DNA, the T-cell specific factor LEF (Giese et al. 1992) augments the activity of positive elements while lacking an intrinsic transcription-activating domain. Alternatively, if FBP can assume multiple conformations, it could possess several distinct DNA-binding specificities.

**FBP, single-stranded DNA, and the regulation of c-myc expression**

FBP is one of an increasing list of proteins that bind tightly and specifically to single-stranded sequences of the human c-myc gene; hnRN protein K (Takimoto et al. 1993) and PUR (Bergemann and Johnson 1992; Bergemann et al. 1992) have also been reported to bind to specific sequences on one strand upstream of the human c-myc gene. Both FBP and hnRN protein K bind to positive cis-acting elements and the activity of these cis-acting elements is augmented by cotransflecting expression vectors for their respective proteins (see above; Takimoto et al. 1993). Significantly, the nucleotides comprising the FUSE site display a pattern of permanganate reactivity in vivo, in which the strand binding FBP in vitro is protected, suggesting the presence of a bound protein, whereas the complementary strand is hyper-reactive, as predicted for naked, single-stranded DNA. These results evoke the possibility that single-stranded DNA-binding proteins play a specific role in the control of c-myc expression. The notion that a sequence-specific single-stranded binding protein can be a transcription factor is supported by the demonstration that CS7.4, a prokaryotic cold shock-specific transcription activator, binds to a specific single-strand sequence (La Teana et al. 1991, Schindelin et al. 1993, Schnuchel et al. 1993).

What roles in gene regulation might best be performed by single-stranded DNA-binding proteins? The alteration in DNA topology and conformation intrinsically linked with transcription suggest several possibilities: Transcription of topologically constrained templates, such as might result from chromosomal attachment to the nuclear matrix, drives positively supercoiled DNA ahead of elongating transcription complexes, whereas negative supercoils accrue behind (Liu and Wang 1987). Considering that RNA polymerase traverses thousands of helical turns in many mammalian transcription units, the potential for topological disruption of normal B-form DNA in active genes is considerable. Torsional stress, accumulating in upstream sequences, can be relieved by unwinding, generating single strands. Stabilization of this underwound DNA with single-strand binding proteins could facilitate the release of transcriptionally generated torsion. Although the formation of Z-DNA, near the FUSE, upstream of actively transcribed c-myc genes has been reported and may provide an additional mechanism to relieve torsional stress (Wittig et al. 1992), at least a nidus of single-strand DNA must still be present as the double helix screws leftward at a B to Z transition. Regulated release of single-strand binding proteins or conversion of single-stranded and Z regions back to B-form DNA could regenerate a negatively supercoiled domain, which might, in turn, facilitate transcription reinitiation, as some mammalian promoters are more active when negatively supercoiled (Parvin and Sharp 1993). In this manner torsional energy is conserved. Alternatively, topoisimerase could dissipate the negative supercoils, in which case the energy stored in the topologically constrained template would not be conserved.

Double-stranded DNA is more rigid than single-stranded DNA. If numerous cis-acting elements are upstream, downstream, and embedded within the c-myc gene, then topological constraints, imposed by loop formation between particular bound proteins and the initiation complexes at the c-myc promoters, may impede the simultaneous or subsequent interaction of the promoter with other bound factors. The introduction of a single-strand bubble could form a flexible joint, facilitating interaction between distant, but topologically constrained, partners.

Single-stranded DNA-binding proteins might also be architecturally suited to bind RNA as well as DNA; the consequences of physiologic competition between these alternative ligands could be exploited to achieve gene regulation. In this manner, the bacteriophage T4 gene 32 single-stranded DNA-binding protein gene autoregulates its own expression by partitioning between mRNA and single-stranded DNA [Lemaire et al. 1978].

Replication is a cellular process associated with the recognition and binding of single-stranded DNA by proteins. Several studies have suggested an origin of replication in the 5'-flanking region of the c-myc gene (McWhinney and Leffak 1990; Vassilev and Johnson 1990). An ensemble of single-strand-specific factors may function directly in replication and, in some manner, exert an indirect influence on c-myc expression. Replication and transcription may also be directly linked. Replication factors acting as activators of transcription for bacteriophage T4 late genes have been demonstrated (Herendeen et al. 1989, 1992), and a similar coupling between replication and transcription in eukaryotes may be implied by the interaction between trans-activation domains and the cellular replication protein RepA [He et al. 1993; Li and Botchan 1993]. FBP could similarly participate in a replication initiation/transcription-activating complex.

Whichever models prove to be correct, the existence of two elements, FUSE and CT, in the c-myc 5' region that stimulate expression augmented by increased levels of proteins shown to bind preferentially to these elements
in single-stranded form suggests that surprising mechanisms may be important for c-myc gene regulation.

Materials and methods

cDNA cloning
Degenerate oligonucleotide pools were synthesized containing sequences that encode the peptides indicated in Figure 2C. The oligonucleotides were used in reverse transcriptase–polymerase chain reactions (RT–PCR) performed as described (Rappolee et al. 1989). In the RT step, 5 μg total U937 RNA, 1.5 μg of random hexamer primers (Promega), and 20 μM of each deoxynucleotide triphosphate (dNTP) were incubated in a 40-μl reaction with Superscript Reverse Transcriptase [GIBCO-BRL] and the manufacturer’s reaction buffer adjusted to 10 mM diethritol at 45°C for 1 hr. After ethanol precipitation and reconstitution in 40 μl of dH2O, 2 μl of the RT reaction products was used as a template in a 50-μl PCR reaction with 20 pmoles each of appropriate primer pairs, 20 nmoles each dNTP, 1.5 units of AmpliTaq polymerase (Perkin-Elmer Cetus), and the manufacturer’s buffer adjusted to 7.5 mM MgCl2. Samples were incubated through 25 cycles of 94°C, 55°C, and 72°C for 2 min at each temperature. PCR products were subcloned into Bluescript(SK-) (Stratagene, La Jolla, CA), and the sequence was determined by dideoxy sequencing [Sequenase, U.S. Biochemical] following the manufacturer’s protocol (Sanger et al. 1977).

The insert from a PCR clone that contained FBP open reading frame was radiolabeled and used to screen cDNA libraries. The three libraries examined were prepared from poly(A)-selected RNA and ligated into the Lambda ZAP II vector (Stratagene). Source RNAs were from undifferentiated HL60 cells, the B-lymphoma cell line BJAB, and PMA/PHA-stimulated human peripheral blood lymphocytes [PBLs]. The complete nucleotide sequence of multiple overlapping clones was determined on both strands with double-stranded plasmid templates by oligonucleotide primer walking. Nucleotide and amino acid sequence analysis was performed with MacVector software (International Biotechnologies). The complete deduced amino acid sequence of FBP was searched against the combined protein sequence data bases at the National Center for Biotechnology Information using the FASTA program. The nucleotide sequence of FBP has been deposited in GenBank under accession number U05040.

RNA blot analysis of FBP transcripts
Polyadenylated RNA was isolated from cultured cells by standard methods [Chirgwin et al. 1979]; samples (10 μg each) were separated on a 1% agarose–formaldehyde gel, transferred to nitrocellulose, and hybridized with a 32P-labeled probe by standard methods [Ausubel et al. 1989]. Blots were washed at a final stringency of 0.2x standard sodium chloride–saturated sodium citrate buffer at 65°C. The FBP probe was a 137 bp PCR fragment containing nucleotides 843–973 of the FBP cDNA, the β-actin probe was a 2.0-kb Psfl fragment from a human β-actin cDNA clone (Gunning et al. 1983). The c-myc probe was a 1.0-kb ClaI–EcoRI fragment from exon 3 [Bentley and Groudine 1986a].

Plasmid construction
The full-length GST–FBP construct was made by joining the sequence of the PBL-1 and PBL-2 cDNAs and ligating the fragment into the Smal site in the pGEX-2T plasmid. To produce the 145–511 construct, the inserts from the HL60-1 and HL60-2 clones were spliced together and the open reading frame region subcloned into the Smal site of the pGEX-2T plasmid. The plasmid that encodes the 278–511 mutant was constructed by inserting an EcoRI–SalI fragment from the HL60-2 clone into the pGEX-1 vector. The fusion protein contains amino acids 278–511 of FBP plus the residues KEIEEKVQE at the carboxy-terminal end stopping at a termination codon unique to the HL60-2 clone. The 278–474-encoding plasmid was constructed from the 278–511 plasmid by inserting a 12-bp double-stranded oligonucleotide with the sequence TTAGTAACTAA into an SfiI site. This oligonucleotide encodes stop codons in all three reading frames so that a truncated protein is produced. The 278–372-encoding plasmid was similarly constructed by inserting the termination oligonucleotide into a DraIII site. The 298–511-encoding plasmid was constructed from the 278–511 plasmid by deleting a fragment between the BamHI site in the pGEX vector and a BclI site in the FBP cDNA.

The insertion mutants were constructed by cutting the 278–511 plasmid at restriction sites, treating the ends with the large fragment of DNA polymerase I [if not already blunt], and inserting linkers of 8, 10, or 12 bp. The appropriate length linker was chosen to insert a small number of amino acids but restore the original reading frame, leaving the rest of the sequence unchanged. In some cases, the blunting and insertion resulted in the elimination of 1 or 2 amino acids at the insertion site. Insertion mutant 1 had the amino acids RIR added between GST and the FBP polypeptide. The remaining insertion mutants had the amino acids listed in Figure 6B inserted or deleted at the positions immediately after the following residues, respectively: mutant 2, M297; mutant 3, C332, mutant 4, A336, mutant 5, L343; mutant 6, P373, mutant 7, Q404, mutant 8, P475. All constructs were confirmed by DNA sequencing.

The eukaryotic reporter plasmids pMP CAT and pMP CATΔ1 are described in Avigan et al. (1990). The pMP CATΔFUSE plasmid was produced from pMP CAT by cutting the parent plasmid at the Aval site located in the Fuse element, partial digestion with T4 DNA polymerase and mung bean exonuclease [GIBCO-BRL] followed by religation. The DNA sequence of the deleted plasmid revealed that 68 nucleotides between position –1493 and –1561 relative to myc P1 were removed, completely deleting the Fuse element. The pMP NsiI–CAT and pMP NsiI–CATΔFUSE reporter plasmids are derivatives of pMP CAT [Avigan et al. 1990] and pMP CATΔFUSE with the deletion of a 580-bp NsiI fragment from position –669 to –1249 relative to the myc P1 promoter. The NsiI deletion results in more consistent CAT expression without disrupting FUSE-mediated activation. The FBP expression plasmids were constructed by inserting the same FBP–cDNA fragment used to produce GST–FBP into the pcDNAI/AMP vector [Invitrogen] in the sense and antisense orientations.

Protein expression, SDS-PAGE, and EMSAs
Recombinant protein was purified from extracts of Escherichia coli [Smith and Johnson 1988] transformed with the pGEX–FBP plasmid constructs on a glutathione–agarose matrix (Sigma). GST alone was prepared from a pGEX–2T plasmid with no insert in a similar manner. Fusion proteins were eluted with 20 mM glutathione, and checked for purity, correct size, and concentration with SDS-PAGE.

EMSAs [Fried and Crothers 1981] were performed with an equivalent amount of purified human or recombinant protein incubated in 25 mM Tris, 200 mM glycine, 1 mM EDTA, 0.5 mg/ml of BSA, 0.1% Tween 20, 10% glycerol, 100 μg/ml of poly[dI–dC], and 0.2 ng of labeled probe. The probe was prepared by 5′-end-labeling synthetic oligonucleotides with T4 polynucleotide kinase in the presence of [γ-32P]-ATP and subsequently...
annealing a complementary strand for double-stranded probes. Reaction mixtures were incubated for 30 min at room temperature, and protein–DNA complexes were resolved by electrophoresis on a 4.2% acrylamide gel in 25 mM Tris, 200 mM glycine, 1 mM EDTA buffer. The FUSE-coding strand oligonucleotide was 5'-GATTCAGAAAATCAGCGAGG-GGGAAAT-3', the noncoding strand, 5'-GATCTATTTTCCCTCATGTTTG-3', encompassing positions -153 to -125 of c-myc. The FUSE noncoding mutant was 5'-GATCTATTTTCCCTCATGTTTG-3' [lowercase indicates changes from FUSE oligomer sequence]. The CTE-coding strand was 5'-AATTCTCTTCCCACCTTCCACCATCCTCCACCTCCCT-3', the noncoding strand, 5'-AAGTTTGCGAGGGTTGGGAAATTGGGAGG-3', encompassing positions -153 to -125 of c-myc.

**Cell culture, transfection, and CAT assays**

HL60 and U937 cells were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum, glutamine, penicillin, and streptomycin. DEAE–dextran transfections were performed as described (Avigan et al. 1990). Electroporations were performed with a Cell-porator (GIBCO-BRL) at 200 V, 1180 μF settings in culture medium. Transfected cells were added to 8 ml of additional medium and incubated for 48 hr before harvesting for CAT assays (Gorman et al. 1982). The same quantity of protein (Bradford method) was assayed for each extract.

**Potassium permanganate treatment of U937 cells and DNA preparation**

U937 cells were sedimented and washed twice in one-tenth culture volume of phosphate-buffered saline at 4°C. Cells were then resuspended in buffer A (15 mM Tris–HCl at pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.5 mM EGTA, 300 mM sucrose, 0.5 mM DTT, 0.1 mM PMSF) to ~3 × 107 cells/ml, split into two tubes, and incubated on ice for 10 min. KMnO4 from a freshly prepared 0.2 M stock was added to a final concentration of 40 mM to one tube and incubated for 30 sec on ice, whereas the second tube received the same volume of TE [for use as a naked DNA control]. Reactions were stopped by the addition of an equal volume of stop solution [50 mM EDTA, 1% SDS, 0.4 M 2-mercaptoethanol], which was also added to the untreated samples. DNA was recovered after two phenol extractions, followed by a phenol/chloroform [1:1] extraction and ethanol precipitation with 400 mM NaCl. The untreated, naked genomic DNA was resuspended subsequently with KMnO4 as described above. Piperidine cleavage was performed as described (Ausubel et al. 1989), and the DNA was resuspended for analysis by LM–PCR.

**LM–PCR**

LM–PCR was performed according the method of Garrity and Wold (1992), with the following modifications. To 5 μl [4 μg] of genomic DNA was added 25 μl of a 1.2× master mix (12 mM Tris–HCl at pH 8.8, 48 mM NaCl, 6 mM MgSO4, 0.012% BSA, 240 μM each dNTP, 2 pmoles of gene specific primer 1, and 0.5 unit of Vent DNA polymerase). Reactions were extracted with phenol/chloroform [1:1], followed by the addition of one-fifth volume of 7.5 m ammonium acetate and ethanol precipitation. Samples were resuspended in 75 μl of ligation solution [50 mM Tris–HCl at pH 7.6, 10 mM MgCl2, 3 mM ATP, 1 mM DTT, 5% PEG-8000, 100 pmoles of unidirectional linker, and 3 units of T4 DNA ligase] and incubated for 4 hr at room temperature. After labeling, samples were extracted with phenol/chloroform [1:1], ethanol precipitated, dried, and resuspended in formamide loading dye. An equal number of cpm of each LM–PCR reaction was separated on a 6% sequencing gel and visualized by autoradiography.

**Acknowledgments**

We are grateful to T. Behrens and L. Staudt for use of the BJAB library and U. Siebenlist for the PBL library, as well as critical review of the manuscript. We wish to thank S. Mackem and M. Lenardo for critical review of the manuscript, P.B. Sigler for helpful discussions, S. Sanford for construction of plasmids, and C. Robinson for technical assistance. M.A. and L.B. were supported by National Institutes of Health grant CA54818.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

**References**


Eick, D. and G.W. Bornkamm. 1986. Transcriptional arrest within the first exon is a fast control mechanism in c-myc


A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif.

R Duncan, L Bazar, G Michelotti, et al.

*Genes Dev.* 1994, 8: Access the most recent version at doi:10.1101/gad.8.4.465

**References**

This article cites 69 articles, 27 of which can be accessed free at: [http://genesdev.cshlp.org/content/8/4/465.full.html#ref-list-1](http://genesdev.cshlp.org/content/8/4/465.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.](#)