The Far Upstream Element-binding Proteins Comprise an Ancient Family of Single-strand DNA-binding Transactivators*

(Received for publication, May 2, 1996, and in revised form, September 4, 1996)

Terri Davis-Smyth†, Robert C. Duncan‡, Tian Zheng, Gregory Michelotti, and David Levens¶
From the Laboratory of Pathology, NCI, National Institutes of Health, Bethesda, Maryland 20892

The cloning and expression of two new human cDNAs encoding proteins highly related to the far upstream element-binding protein (FBP) are described. FBP, FBP2, and FBP3 comprise a family of single-strand DNA-binding proteins that possess all of the general features of more conventional transcription factors. The FBP's each bind sequence specifically to only one strand of the far upstream element (FUSE; originally identified upstream of c-myc), and each possesses potent activation domains when fused to the GAL4 DNA-binding domain and assayed by transient transfection. Typical of transcription factors, the proteins are most highly related in their central, DNA-binding domains, but extensive homology is also shared within the tyrosine-rich, carboxy-terminal activation domains. Comparison with GenBank sequences revealed a fourth FBP family member encoded by Caenorhabditis elegans chromosome III, illustrating the high degree of homology in this evolutionarily ancient and conserved family.

Most sequence-specific DNA-binding proteins use cis elements in B-form DNA as platforms from which effector domains are thrust toward the transcriptional apparatus (1, 2). Less well characterized and more problematic are proteins docking onto DNA sites in non-B conformation (3–8). The intracellular, physiological conditions promoting non-B-form DNA, permitting single-strand or Z-DNA-specific proteins to bind, have not been defined; the linkage between these unusual proteins and specific gene regulation has been obscure. Recently, we described a cDNA clone and characterized the encoded protein that binds to a specific single-strand sequence far upstream of the human c-myc gene (9). This far upstream element (FUSE)ψ is most active in undifferentiated cells and becomes inactive as differentiation is induced and transcription initiation is shut off (10). In parallel, this single-strand FUSE-binding protein (FBP) is present in undifferentiated cells and is down-regulated with differentiation (9, 11). Expression of recombinant FBP in leukemia cells augments the positive activity of the FUSE element (9). The FBP is composed of three domains: amino-terminal, central, and carboxy-terminal. The central domain, consisting of four copies of a reiterated unit comprised of a 30-residue direct repeat and an amphipathic helix separated by 18 to 21 residues, specifies structures sufficient for single-strand DNA binding and sequence recognition. The third and fourth copies of this repeat-helix unit constitute a minimum sequence-specific, single-strand DNA-binding domain (9). Although this pattern of repeats and helices does not resemble any known DNA-binding motif, the repeats alone are similar to the K homology (KH) motif recurrent in hnRNP K (12, 13). Although hnRNP K was originally identified as a component of hnRNP complexes, it has subsequently been shown to be a sequence-specific single-strand DNA-binding protein, preferring DNA to RNA and possessing a transactivation domain (7, 8). The carboxyl terminus of FBP displays an unusual repeated sequence. Tyrosine-rich and possessing powerful trans-activating properties, this sequence can be transferred to heterologous B-form DNA-binding domains, transforming them into potent chimeric transcription factors (14).

The evolutionary expansion and diversification of transcription factors such as heat shock factors, nuclear factor-κB, homeobox genes, and steroid receptors into families and superfamilies is proof of their physiological utility (15–20). Are FBP and hnRNP K molecular quirks as gene regulators, acting through peculiar single-stranded or supercoiled elements, or might they be representative of a larger class of molecules sensitive to the conformation and/or topology of their target sequences? If the strategies exploited by FBP were of general consequence, then the structural features of FBP governing its interaction with DNA might be both ancient and diverse. Filling these predictions, we describe the cDNAs encoding and characterization of two new FBP-like molecules and report remarkable conservation between the human FBP family and a Caenorhabditis elegans cousin. Remarkably, the reiterated architecture, including the inter-repeat distance, is preserved in this family, as is the short activating sequence.

MATERIALS AND METHODS

Library Screening and cDNA Sequencing—A random and oligo(dT)-primed human skeletal muscle cDNA library (Clontech) was probed with 32P-labeled human FBP cDNA encoding the third and fourth repeats (9) under standard conditions (40% formamide, 4 × SSC, 10% dextran sulfate, 0.1% SDS, and 20 μg/ml sonicated salmon sperm at 42 °C). Filters were washed at reduced stringency (6 × SSC, 0.1% SDS at room temperature). The oligo(dT)-primed Raji B-cell cDNA library was a generous gift from L. Staudt. Hybridization conditions were similar to that above, but the wash stringency was increased (0.2 × SSC, 0.1% SDS at 42 °C).

Sequencing was determined by dideoxy sequencing using Sequenase version 2.0 (U. S. Biochemical). Nucleotide and amino acid sequence analysis was performed with MacVector software (International Biotechnologies, Inc.).

Tissue Culture and Northern Blot Analysis—U937 cells were main-
FIG. 1. Nucleotide sequences and predicted amino acid sequences of FBP2 (A) and FBP3 (B) cDNA clones. Nucleotides are numbered above each line. The termination codons are indicated by an \( \text{p} \). Not annotated are poly(A) addition sites at 2403 in FBP2 and 3104 in FBP3. Conservation and homology between the FBP proteins. The three domains: amino-terminal, central (also the DNA-binding domain), and carboxyl-terminal are noted. The structural features of the proteins are represented schematically. GGG signifies the glycine-rich segments within the amino-terminal; the solid boxes depict the KH repeats in the central domain; the wavy-line filled boxes symbolize the amphipathic helices; and the tyrosine-rich section of the carboxyl-terminal is depicted as AW/A/EY. The percent homology of each region of FBP2 and FBP3 compared with FBP is indicated below the corresponding domain. The segment encoded by the original FBP cDNA probe used to screen the human skeletal muscle library is indicated by a line above the FBP schematic.
tained in RPMI 1640 with 10% fetal calf serum (v/v). Mid-log cells were sedimend and resuspended at 250,000 cells/ml in the same media containing 50 ng/ml 12-0-tetradecanoylphorbol-13-acetate, which was prepared from a 200 mg/ml stock in dimethyl sulfoxide. Control cells treated with dimethyl sulfoxide alone did not differentiate, as assayed by c-myc Northern blot analysis. Cells were divided into six aliquots of $1.9 \times 10^7$ cells. After 0, 8, 24, 48, 72, or 96 hours of 12-0-tetradecanoylphorbol-13-acetate treatment, the cells were harvested.

**FIG. 1—continued**
Total cytoplasmic RNA was isolated using RNazol B (Tel-Test, Inc.). Standard Northern blot analysis was performed as described previously (21). RNA was separated on 1.2% agarose gel containing 6% formaldehyde and blotted onto nitrocellulose. Hybridization was carried out in 40% formamide, 10% dextran sulfate, 0.1% SDS at 42°C overnight. The filter was washed in 0.2× SSC, 0.1% SDS at 65°C for 30 min and then exposed to film (Kodak XAR-2).

**Plasmid Constructs**—Glutathione S-transferase (GST) fusion proteins were constructed by cloning each of the FBPs in frame into a pGEX vector (AMRAD Corp. Ltd.). The full-length GST-FBP construct was cloned and expressed as described previously (9, 22). The GST-FBP2 and GST-FBP3 fusion proteins were similarly constructed and expressed.

Expression plasmids fused the carboxyl-terminal domain of FBP, FBP2, and FBP3, generated by polymerase chain reaction from the respective cDNA to the GAL4 DNA-binding domain (14). G4/FBPC contains amino acids 448 through to the end of the C terminus, G4/FBP2C possesses residues 439 through its terminus, and G4/FBP3C also has residues 439 through to the end. The G4/E1a plasmid fuses residues 121 to 223 of the adenovirus E1a protein to GAL4(1–147) as described previously (23). The reporter plasmid containing five GAL4-binding sites upstream of a minimal E1b TATA sequence and the CAT coding sequence were also described previously (23).

**Electrophoretic Mobility Shift Assays, Transfections, and CAT Assays**—Electrophoretic mobility shift assays were performed with recombinant GST fusion proteins, 4 ng of FBP, 16 ng of FBP2, and 12 ng of FBP3. The varying quantities of protein bound equivalent amounts of FUSE wild-type probe. The recombinant protein was incubated in 25 mM Hepes, 100 mM NaCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride, 0.5% Tween 20, 20% bovine serum albumin, 20 mg/ml poly[d(I-C)], and 6 fmol of 5’ end-labeled probe. Where indicated, competitor oligonucleotides were added. The reaction was incubated for 30 min at room temperature, and the protein-DNA complexes were resolved by electrophoresis on a 4% native acrylamide gel in Tris-borate EDTA. The complexes were visualized by autoradiography.

HeLa cells were cultured in Dulbecco’s modified Eagle’s medium (Life Technologies, Inc.) with 10% fetal calf serum. Transfections were performed via electroporation as described by Tomonaga and Levens (8). Following 48 h of posttransfection culture, cell extracts were prepared, normalized for protein content, and assayed for CAT activity as described previously (21, 24).
TABLE I
Codon usage by the FBPs

| Amo- | Codons most frequently used to specify human protein | Codons least frequently used to specify human protein |
|------|----------------------------------------------------|--------------------------------------------------|
| FBP  | 36% (234/644)                                      | 58% (371/644)                                     |
| FBP2 | 66% (428/651)                                      | 29% (190/651)                                     |
| FBP3 | 56% (338/600)                                      | 37% (220/600)                                     |

RESULTS

Cloning of FBP-related cDNAs—A cDNA fragment of FBP, encoding repeat units 3 and 4 (Fig. 1C), was labeled and hybridized to a multi-tissue blot of human RNA; the presence of several faint but discrete transcripts in skeletal muscle and brain, migrating differently than FBP mRNA, justified the use of this same probe and similar hybridization conditions to screen a human skeletal muscle library (data not shown). In addition to six clones of FBP, this screen yielded three clones constituting overlapping segments of a sequence highly related to, but distinct from, FBP. The longest of these clones, 1350 nucleotides, specified a protein possessing four FBP-like repeats but lacked initiator methionine or termination codons. After the skeletal muscle library failed to yield additional independent clones for this FBP homolog, a Raji B-cell line library was screened under moderately stringent conditions using the new FBP homolog as a probe. This library has yielded long cDNAs for several other genes.) This screen yielded two particularly noteworthy clones. The sequence of the first clone, 2445 bases in length (Fig. 1A), totally encompassed the 1350-base protein coding segment of the first FBP-like clone; throughout the open reading frame, the predicted protein was highly related to FBP. We designated this protein FBP2; several other independent clones encoded shorter segments of FBP2. A second unique cDNA, 3169 nucleotides in length (Fig. 1B), specified a different FBP-like protein and was named FBP3.

Properties of FBP2 cDNA Sequence and Predicted Protein—The open reading frame of the FBP2 clone terminates after 1957 bases; the 3'-untranslated region specifies a poly(A) addition signal at residue 2403, and the poly(A) tract commences at position 2422. The codon usage between FBP and FBP2 are startlingly different, belying their strong homology. Whereas FBP1 relies heavily on codons not frequently used to specify human protein, FBP2 is coded primarily by the most frequent codons (Table I) (25). An initiator methionine codon and translation start site is lacking; despite numerous attempts, applying a variety of methods, additional 5' FBP2 cDNA sequence has not been obtained.

The cDNA clone of FBP2 encodes an amino acid sequence of 652 residues. A Pustell matrix comparison of the deduced translation of FBP2 with FBP (Fig. 2A) revealed the same general architecture found in both proteins. Three distinct domains in FBP2, amino-terminal, central, and carboxy-terminal, displayed strong primary sequence and predicted secondary structure homology with FBP. The amino- and carboxyl-terminal domains each shared approximately 60% homology with FBP, whereas the central nucleic acid-binding domains shared greater than 80% homology (Fig. 1C). Notably, the central nucleic acid-binding domain possesses four regularly spaced hnRNP K homology motifs, each followed by a predicted amphipathic helix, with spacing virtually identical to that within FBP.

Properties of FBP3 cDNA Sequence and Predicted Protein—The 3169 nucleotides of the FBP3 cDNA encode 600 amino acids ending with TAG at nucleotide 1809. A polyadenylation signal at 3104 is followed by a poly(A) tract starting at residue 3140. The open reading frame of FBP3 is intermediate between FBP and FBP2 with respect to preferential utilization of uncommon codons.

A Pustell matrix comparison of the deduced translation product of FBP3 with FBP (Fig. 2B) once again revealed preservation of the same general architecture in both sequences. The same three distinct domains, amino-terminal, central, and car-
boxyl-terminal, were apparent and shared strong primary sequence and predicted secondary structure homology with FBP. The amino- and carboxyl-terminal domains were slightly more divergent from FBP than were the equivalent segments of FBP2, but both FBP2 and FBP3 were similarly homologous with FBP within their central domains (Fig. 1C). As with FBP2, the four hnRNP K homology segments with interspersed amphipathic helices are precisely spaced.

FBP, FBP2, and FBP3 Are Widely Expressed but Distinctly Regulated—No clear rule governs the expression patterns of the individual members of protein families. The homologous members of some families, e.g., the c-myc family (c-myc, N-myc, and L-myc), display distinct tissue-restricted patterns of expression, distinguishable from each other under physiological conditions; thus, pathological deregulation of the different myc proteins are associated with tumors of different origin (26, 27). In contrast, the jun family of homologous oncogenic transcription factors are often expressed within the same cell but display different transcriptional or posttranscriptional regulatory profiles (21, 28, 29). Are each of the FBPs functionally equivalent but expressed in different cell types, or are they expressed within the same cell as components of distinct or overlapping circuits? FBP mRNA has been shown to be expressed broadly in different tissues and cell lines but to varying levels (9, 11). It is most highly expressed in dividing cells with a temporal and regulatory profile paralleling that of c-myc (9, 11). To determine if different FBPs might be associated with different cell types, perhaps mirroring the myc, probes selective for FBP, FBP2, or FBP3 were prepared and hybridized at high stringency to RNA extracted from a variety of cells and tissues. As reported previously, FBP probes hybridized with a 2.6-kilobase transcript; in contrast, probes for both FBP2 and FBP3 annealed with mRNAs approximately 4.5 and 3.4 kilobases, respectively. All of the FBPs were expressed in all of the cells and tissues examined but to varying absolute and relative amounts (Fig. 3, lanes 1–10). In different cells and tissues, one or another FBP family member was most abundant. No simple pattern of tissue or cell specificity was observed for each individual FBP or for the FBP family in aggregate. As reported previously and confirmed here, down-regulation of the 2.6-kilobase FBP mRNA commenced in growing cells soon after the induction of differentiation in the promonocytic leukemia cell line U937 (Fig. 3, lanes 11–16). In contrast, FBP3 mRNA (Fig. 3, lanes 11–16) displayed no distinct shut off upon induction of differentiation and FBP2 mRNA shut off was incomplete, with levels gradually dwindling after 3 days. Similar shut off of FBP with lingering expression of FBP2 and FBP3 was also seen with the promonomyelocytic leukemia cell line HL60 (data not shown). Whereas FBP expression generally parallels the excursions in the amount of c-myc mRNA present in a cell, the patterns of FBP2 and FBP3 transcription fail to evoke comparison with immediate-early, growth-regulated gene expression.

As an additional note, adding a presumed 300-residue poly(A) tract to the cDNA sequences of FBP2 and FBP3 accounts for 2721 and 3440 nucleotides of their transcripts. As both FBP2 and FBP3 sequences extended 5′ from the poly(A) tail through open reading frames predicting proteins contiguously homologous with the full extent of FBP, then both FBP2 and FBP3 proteins should possess amino termini extending beyond that of their close relative, FBP. Although almost all of FBP3 mRNA is accounted for by the available sequence, some residues are likely to be missing from the amino terminus because no initiator ATG is found. A larger portion of the 5′ end of the FBP2 cDNA is lacking. An 80-kDa nucleic-acid-binding protein from HeLa cells has been identified that cross-reacted with antibodies against FBP and yielded two tryptic peptides displaying sequence identity with FBP2; therefore, it is possible that approximately 300 additional coding nucleotides occur 5′ of the present FBP2 open reading frame. Repeated attempts to extend the 5′ sequences of FBP2 and FBP3 through screening multiple libraries, 5′ rapid amplification of cDNA ends, and genomic sequencing all failed to yield additional information in this regard or to provide enough information to prepare additional primers to extend the present sequence. It is likely that the bulk of the missing nucleotides of the FBP2 mRNA comprise 5′-untranslated region rather than open reading frame.

Protein Sequence and Conservation of Domains among the FBPs—Optimal alignment between the three human FBPs reveals numerous conserved regions (CRs I–XII) as annotated in Fig. 4A. Recently, sequence from C. elegans chromosome III (CELZK418_9: gP3U00047) was predicted to encode a protein sharing many features with the FBPs as is illustrated by the Pustell matrix comparison shown in Fig. 2C. Most of the sequences conserved between the human FBPs correspond with proposed structural elements first noted in FBP. The immediate amino-terminal segment of all the FBPs is glycine-rich (CR I) and is followed by CR II, previously proposed to form an amphipathic helix in FBP (9). The distribution of hydrophilic and hydrophobic residues within this helical segment is com-

---

A. MacMillan and P. A. Sharp, personal communication.
pletely conserved between family members. The function of conserved region III has not been defined, but conservation of this element extends to the amino terminus of *C. elegans* FBP, which encodes GIGTLKRPLDSEI, most closely resembling that found in FBP3.

CRs IV–VIII comprise the most highly conserved region of the FBP family and for FBP constitute a sequence-specific DNA-binding motif. Each of these conserved segments contains a primary sequence element sometimes referred to as a KH motif. KH motifs are found within a variety of proteins capable of interacting with DNA or RNA (12, 13, 30–32). The KH repeats of each of the FBPs (including the *C. elegans* homolog) are shown in Fig. 4B. Also shown, for comparison, are the KH repeats found in hnRNP K, which also binds to a single-stranded element of the human *c-myc* gene (7, 8). The notion that KH repeats are important for sequence-specific DNA binding is supported by biochemical studies of deletion mutants of FBP and hnRNP K (8, 9). That KH motifs mediate sequence recognition in DNA is supported by the crystal structure of *Hha*I methylase; a segment of that protein contacts and distorts the DNA helix upon binding which includes a sequence indistinguishable from a KH motif (33, 34). The stereotypical pattern of the FBP family central domain displays four regularly spaced, alternating units, each consisting of a KH motif and an amphipathic helix. The conservation of the pattern, spacing, and sequences of the repeating units, extending from *C. elegans* to the human FBPs, argues that the entire central domain has evolved for a particular function and is not an amalgam of independent nucleic acid recognition motifs.

The carboxyl terminus of FBP, which possesses a potent transcription activating domain, bears four conserved segments (14). Although the function of CR IX is not known, CRs X–XII are variants of a tyrosine-rich, transcription-activating motif (14). In FBP, CRs X–XII each occur once, whereas in FBP2, CR XI is duplicated, and in FBP3, CR X is deleted. *C. elegans* possesses a single variant of CR XII but additionally displays a potential glutamine-rich activating sequence within its carboxyl-terminal segment (35).

Comparison of the amino acid sequences of FBP, FBP2, and FBP3 in regions where all three proteins overlap reveals that FBP2 and FBP3 are more dissimilar from each other than either is with FBP, indicating that FBP is likely to be the oldest member of the family. FBP is slightly more related to *C. elegans* FBP than are FBP2 or FBP3.

**The FBP Family of Single-strand Binding Transactivators**

The degree of primary structure conservation shared between the nucleic acid-recognizing, central domains of the FBPs suggested that all three family members would interact similarly with DNA. To test this prediction, GST was fused to the central domains, including all four repeating units of FBP, FBP2, or FBP3; chimeric fusion proteins were purified and tested for specific binding to the human *c-myc* FUSE sequence. As shown previously, FBP bound to the FUSE specified by the template strand of the human *c-myc* gene (9, 10). This binding was specific and not eliminated by inclusion of heterologous competitor deoxyoligonucleotide or poly[d(I-C)]. A series of mutant FUSE oligonucleotides was generated by systematically altering blocks of four residues throughout the probe sequence; the mutations introduced were all transversions (Fig. 5B). As expected, some of the mutations perturbed the interaction of FBP with FUSE, as demonstrated by lack of competition (Fig. 5A, lanes 5 and 7), whereas other mutations did not impair FBP binding and served as effective competitors (Fig. 5A, lanes 6, 8, 9, 10, and 11). GST-FBP2 or GST-FBP3 was incubated with FUSE probe and challenged with the same series of competitor oligonucleotides described above. Both FBP2 and FBP3 displayed analogous binding properties, indicating that all three family members would recognize similar DNA sequences.
played binding profiles that closely paralleled that of FBP. In addition, both FBP2 and FBP3 bound neither to double-stranded FUSE nor to single-stranded FUSE sequence from the non-template strand of human c-myc (data not shown).

The FBPs Activate Gene Expression—FBP was originally identified as a factor binding to FUSE, a positive cis-element of the human c-myc gene (9, 10). Subsequently, recombinant FBP was shown to augment c-myc promoter-mediated expression in transient transfection assays (9). More recently, the carboxyl terminus of FBP has been grafted onto the GAL4 DNA-binding domain, yielding a potent GAL4-UAS transactivator (14). The individual tyrosine-rich motifs are sufficient to activate reporter expression, and the tyrosines are, in fact, necessary for this activity (14). The conservation of tyrosine motifs in FBP2 and FBP3 relative to FBP was sufficiently strong to invite similar experiments, testing whether the carboxyl termini of these proteins could also activate gene expression. FBP, FBP2, and FBP3 coding sequences immediately downstream of the central domain and extending past the termination codons were fused in frame to the GAL4 DNA-binding domain and cotransfected with a GAL4-UAS-driven reporter. The carboxyl termini of all of the FBP3s proved to be potent activators of gene expression (Fig. 6). No clear correlation between the number of tyrosine-rich repeats and the degree of transactivation was apparent.

DISCUSSION

FBP, FBP2, and FBP3 define a family of homologous gene-regulatory proteins capable of tethering a powerful activation motif to specific sequences in single-stranded DNA. Conceptual splicing and translation of a gene found on C. elegans chromosome III identifies a fourth member of this family and attests to the ancient origin of these proteins. The expression of FBP, the prototype of this family, parallels c-myc expression under a variety of conditions, implicating FBP as a candidate for regulating cell growth and differentiation (9, 11). In contrast, no simple pattern of tissue-specific, differentiation-specific, or growth-regulated expression of FBP2 or FBP3 has yet been discerned. At the extremes, the FBP3s may concertedly cross-regulate the same downstream targets or, alternatively, in in vivo conditions restrict the binding and trans-activating specificities of the FBP3s sufficiently to target their actions to different genes. The spectrum of genes regulated by the FBP3s remains to be defined. Although one of these genes is the c-myc proto-oncogene, other targets certainly exist. Because recognizable myc proteins first appear only with vertebrates, it is apparent that the FBP3s, clearly recognizable in invertebrates, predates its downstream target c-myc. Most likely, the evolving c-myc gene co-opted existing regulatory components, including FBP, to affect its proper regulation.

The minimal molecular characteristics defining trans-factors regulating gene expression include: specific cis-element recognition, usually constituted within a single domain; and effector function, most frequently residing in a separate domain, although not necessarily independent of the former (2, 15, 23, 34–39). Although all of the FBP3s possess both specific nucleic acid recognition and effector domains, the molecular architecture of these domains are unusual. The nucleic acid-recognizing central domain of each FBP displays a reiterated array of repeating units. Part of each unit is a primary structure element dubbed the KH motif, so designated because it was first
recognized within the primary structure of hnRNPK (12). Although the KH motif is commonly believed to be an RNA-recognition scaffold, it is actually a single-strand nucleic acid recognition platform (8, 12, 13). hnRNPK, the archetypal KH protein, when challenged with its cognate cis-element either as single-stranded DNA or RNA, strongly prefers DNA (8). Furthermore, other KH proteins, such as p62 and FMR, form stable complexes both with DNA and RNA (29, 31). Most bona fide RNA-binding proteins can discriminate against DNA quite well under competitive conditions; the dual specificity of KH proteins invites further study.

All of the FBPs bind specifically to single-stranded DNA. Under what physiological conditions might duplex cis-elements, such as FUSE, become sufficiently destabilized to allow binding of the FBPs? Replication, recombination, and repair all expose single-stranded regions. In fact, several reports suggest that an origin of replication occurs within the vicinity of the FUSE element (4). Binding of the FBPs to melted FUSE, exposed at or near a replication origin, provides a potential device to couple c-myc expression with DNA synthesis. Transcription may expose FBP-binding sites. Obviously, as elongation complexes transit the body of a primary transcript, the template is obligatorily unwound. Furthermore, recent experiments have indicated that FBP and hnRNPK target their cognate sequences in negatively supercoiled regions; the destabilization of B-DNA in these torsionally strained regions is apparently sufficient to allow FBP to bind (6). Similar destabilization of binding sites might occur as transcription elongation generates and transmits negative supercoils to upstream elements in a topologically constrained domain. Thus, factors that recognize topological strain conceptually might act as real time sensors of promoter activity.

The activation domain of the FBPs is unusual. Known activation domains are frequently so enriched for particular amino acids as to appear, on occasion, virtually homopolymeric. Acidic residues, glutamine, proline, glycine, and isoleucine-rich activation domains have all been described (35–38). The FBPs all possess two or four repeats of a tyrosine-rich activation motif. Replacement of the tyrosines by phenylalanine greatly impairs activation, indicating a role for the phenolic hydroxyl in FBP function. The paired tyrosine dyads found in all of the FBPs resemble known substrates of tyrosine kinases. Modulation of activation potential through direct phosphorylation of tyrosine residues in the activation domain is an intriguing possibility (14, 39). If this scenario were to play out, then the different FBPs could potentially respond to different signals, depending on the exact sequence of each family member and the exact substrate specificity of the relevant kinases.

The physiological processes requiring or facilitated by multiple FBPs remain obscure. Little is known of the nature or number of genes that might be responsive to the FBPs. Although all of the FBP DNA-binding domains interact with the FUSE of c-myc, this observation is no more than a coarse indicator of the nucleic acid recognition properties of these molecules. Subtle differences in amino acid sequence may alter the spectrum of sequences and genes recognized and regulated by each protein. For example, the nuclear factor-x-B-rel family of proteins possess distinct and finer specificity for DNA binding and physiological action than is discernible from binding to the immunoglobulin κ intron enhancer. Even now, the enumeration and elucidation of the target genes and physiological roles of each nuclear factor-x-B-rel family homo- or heterodimers are incomplete. Likewise, the structural similarity of GATA-1, GATA-2, and GATA-3 belies distinctions in their target sequences, tissue distribution, and physiology (40, 41). The FBPs are unlikely to be redundant proteins dedicated solely to c-myc control. Direct and indirect evidence indicates that the FBPs respond to different signals. The most clear evidence of distinct physiology is the rapid shutoff of FBP in contrast to the lingering expression of FBP2 and FBP3 during differentiation of leukemia cells. The peculiar utilization of rare codons to specify FBP versus the specification by frequent codons for FBP2 may foreshadow some form of differential translational regulation reflecting different physiology for these proteins. The unusual features of the FBPs may highlight interconnections between signal transduction, RNA synthesis, DNA topology, and DNA conformation.

Acknowledgments—We thank Suzanne Sanford for sequencing and synthesis of oligonucleotides, Charles Robinson for technical assistance and Takeshi Tomonaka for RNA prepreations.

REFERENCES
1. Ptashne, M. (1988) Nature 333, 683–689
2. Tjian, R., and Maniatis, T. (1994) Cell 77, 5–8
3. Aliti, S., and Groener, B. (1993) Mol. Cell. Biol. 13, 7303–7310
4. Bergemann, A. D., and Johnson, E. M. (1992) Mol. Cell. Biol. 12, 1257–1265
5. Chang, L. L., Clawson, M. H., Bilgrami, S., and Carmichael, G. (1995) Cell Growth Diff. 4, 1975–1983
6. Michieltotti, G., Michieltotti, E. F., Pulliner, A., Duncan, R. C., Eick, D., and Levens, D. (1996) Mol. Cell. Biol. 16, 2856–2869
7. Takimoto, M., Tomonaka, T., Matumis, M., Avigan, M., Kutzhcz, H., Dreyfuss, G., and Levens, D. (1993) J. Biol. Chem. 268, 18249–18258
8. Tomonaka, T., and Levens, D. (1995) J. Biol. Chem. 270, 4875–4881
9. Duncan, R. D., Bazar, L., Michieltotti, G., Tomonaka, T., Kutzhcz, H., Avigan, M., and Levens, D. (1994) Genes Dev. 8, 465–480
10. Avigan, M. I., Struber, B., and Levens, D. (1990) J. Biol. Chem. 265, 18538–18545
11. Bazar, L., Harris, V., Sunithi, I., Hartmann, D., and Avigan, M. I. (1995) Oncogene 10, 2229–2238
12. Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L., and Dreyfuss, G. (1994) Cell 77, 33–39
13. Siomi, H., Matumis, M. J., Michael, W. M., and Dreyfuss, G. (1993) Nucleic Acids Res. 21, 1193–1196
14. Duncan, R. D., Collins, I., Tomonaka, T., Zheng, T., and Levens, D. (1996) Mol. Cell. Biol. 16, 2274–2282
15. Evans, R. M. (1988) Science 240, 889–895
16. Gehring, W. R., Affolter, M., and Burghlin, T. (1994) Annu. Rev. Biochem. 63, 457–526
17. Nakai, A., and Morimoto, R. I. (1993) Mol. Cell. Biol. 13, 1983–1997
18. Ruddle, F. H., Bartels, J. L., Bentley, K. L., Krappon, C., Murtha, M. T., and Pendleton, J. W. (1994) Annu. Rev. Genet. 28, 425–445
19. Siebenlist, U., Fancoso, G., and Brown, K. (1994) Annu. Rev. Cell Biol. 10, 405–455
20. Tsai, M. J., and O’Malley, B. W. (1994) Annu. Rev. Biochem. 63, 451–486
21. Faull, A. R., Davis-Smyth, T., Gardner, K., and Levens, D. (1995) J. Biol. Chem. 269, 2646–26475
22. Smith, D. B., and Johnson, K. S. (1988) Gene (Amst.) 76, 31–40
23. Lillie, J. W., and Green, M. R. (1989) Nature 338, 39–44
24. Gorman, C. M., Moffat, L. F., and Howard, B. (1982) Mol. Cell. Biol. 2, 1044–1051
25. Wada, K., Wada, Y., Ishibashi, F., Gojobori, T., and Ikemura, T. (1992) Nucleic Acids Res. (Suppl.) 20, 2111–2118
26. Garte, S. J. (1993) Crit. Rev. Oncogene 4, 435–449
27. Prins, J., DeVries, E. G., and Mulder, N. H. (1993) Anticancer Res. 13, 1373–1385
28. Kovary, R., and Bravo, R. (1991) Mol. Cell. Biol. 11, 2451–2459
29. Ryseck, R. P., and Bravo, R. (1990) Oncogene 6, 533–542
30. Ashley, C., Wilkinson, K., Reines, D., and Warren, S. (1993) Science 262, 563–566
31. Gibson, T. J., Thompson, J. D., and Heringa, J. (1993) FEBS Lett. 324, 361–366
32. Wang, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakas, P., and McCormick, F. (1992) Cell 69, 551–558
33. Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994) Cell 76, 357–369
34. Roberts, R. J. (1995) Cell 82, 9–12
35. Courrey, A., and Tjian, R. (1988) Cell 55, 887–889
36. Attardi, G. D., and Tjian, R. (1983) Genes Dev. 7, 1341–1353
37. Mermel, N., O’Neill, E., Kelly, T., and Tjian, R. (1989) Cell 58, 741–753
38. Meijer, D., Graus, A., and Grosveld, G. (1992) Nucleic Acids Res. 20, 2241–2247
39. Holloway, J. M., Saito, D. P., Scully, K. M., Glass, C. K., and Rosenfeld, M. G. (1995) Genes Dev. 9, 1992–2006
40. Merika, M., and Orkin, S. H. (1993) Mol. Cell. Biol. 13, 3999–4010
41. Ko, L. J., and Engel, J. D. (1993) Mol. Cell. Biol. 13, 4011–4022