Isolation of a recombinant copy of the gene encoding C/EBP

William H. Landschulz, Peter F. Johnson, Eli Y. Adashi,[1] Barbara J. Graves,[2] and Steven L. McKnight

Department of Embryology, Carnegie Institution of Washington, Baltimore, Maryland 21210 USA

In two previous studies we described the properties of a heat-stable DNA-binding protein present in rat liver nuclei. This protein, hereafter termed C/EBP, is capable of selective binding to the CCAAT homology of several viral promoters (Graves et al. 1986), as well as the core homology common to many viral enhancers (Johnson et al. 1987). We now report the isolation of a recombinant clone of the gene that encodes C/EBP. Expression of the clone in bacterial cells yields a protein that binds in vitro to both the CCAAT homology and the enhancer core homology, providing conclusive evidence that a single gene product accounts for both binding activities. By examining the properties of protease-derived fragments of C/EBP, we have localized its DNA-binding domain to a 14-kD fragment. A 60-amino-acid segment located within the DNA-binding domain of C/EBP bears sequence similarity to the products of the *myc* and *fos* oncogenes.

The expression of mRNA-coding genes in eukaryotic cells can be modulated by *cis*-regulatory DNA sequences that are located thousands of base pairs away from the site at which RNA polymerase II begins transcription. DNA sequences that can exert a positive influence on transcription from a long distance have been termed enhancers. Enhancers play an integral role in the regulation of both viral and cellular gene expression (e.g., Banerji et al. 1981, 1983; Moreau et al. 1981; Fromm and Berg 1983; Gillies et al. 1983; Queen and Baltimore 1983; Walker et al. 1983).

Molecular genetic studies have indicated that enhancers are composed of an amalgamation of independent *cis*-regulatory motifs (Herr and Clark 1986; Zenke et al. 1986; Schirm et al. 1987). Crude nuclear extracts prepared from cells and tissues have been found to contain proteins that bind to these *cis*-regulatory motifs (e.g., Piette et al. 1985; Wildeman et al. 1986; Speck and Baltimore 1987), and, in several cases, substantial progress has been made toward the isolation and purification of these factors (e.g., Treisman 1986; Prywes and Roeder 1986; Lee et al. 1987; Johnson et al. 1987; Mitchell et al. 1987).

The findings of studies such as those outlined in the preceding paragraphs have established a useful framework for further study. Each *cis*-regulatory motif of an enhancer interacts with a specific *trans*-acting factor. As such, the activity of an enhancer is a function dictated both by its array of *cis* motifs and the availability of factors that bind to those motifs. Despite its conceptually satisfying nature, this model fails to address one of the more interesting and enigmatic aspects of enhancer action. That is, how can an enhancer influence the transcription of a gene that is located thousands of base pairs away?

Our perspective on the mechanistic basis of enhancer action has been influenced by two additional observations. First, Maniatis and colleagues have shown that mutations that subtly alter the promoter of a target gene can block enhancer action (Treisman et al. 1983). Apparently, enhancers act to elevate transcription by 'communicating' with promoter-associated *cis*-regulatory motifs. Second, work on a variety of different prokaryotic regulatory proteins has shown that multimeric forms of a single polypeptide can facilitate the juxtaposition of binding sites on DNA that are separated by hundreds of nucleotides (Dunn et al. 1984; Dandanell and Hammer 1985; Hochschild and Ptashne 1986; Reitzer and Magasanic 1986; Kraemer et al. 1987).

We have discovered a DNA-binding protein, termed C/EBP, that exhibits properties that might be expected of a molecule capable of facilitating communication between enhancers and promoters. The two *cis*-regulatory motifs recognized by C/EBP are the CCAAT homology (5'-CCAAT-3'), which is common to many promoters, and the enhancer core homology (5'-TGTGG$^A_T$$^A_T$$^A_T$G-3'), which is common to many enhancers. Aside from its ability to recognize two *cis*-regulatory motifs that bear no overt DNA sequence similarity (CCAAT homology and enhancer core homology), C/EBP exhibits a tenacious capacity to retain DNA-binding activity following

either thermal or chemical denaturation. To study these unusual properties of C/EBP, we have isolated a cDNA copy of its mRNA, and developed an antiserum that recognizes C/EBP in its native form.

Results

Derivation of partial amino acid sequence and antibodies to C/EBP

Standard procedures of protein fractionation can be used to purify a 20-kD protein from freshly dissected rat liver that binds to the both the CCAAT homology and the enhancer core homology (Johnson et al. 1987). We are unaware of the specific regulatory role played by this protein in rat liver or other cell types. As such, we designate it only according to the binding specificities that we have observed in vitro (CCAAT/Enhancer Binding Protein = C/EBP). Several attempts to obtain partial amino-terminal amino acid sequence from the 20-kD form of C/EBP failed. When frozen rat livers were substituted for fresh material, our fractionation scheme led to the purification of a series of polypeptides that were 4–10 kD shorter than the initially characterized form of C/EBP (data not shown). Four of these smaller polypeptides were isolated as homogeneous species following excision and electroelution from an SDS-polyacrylamide electrophoresis gel. After renaturation through 6 M guanidine hydrochloride (Hager and Burgess 1980), each polypeptide was tested for sequence-specific DNA-binding activity by direct DNase I footprinting, and was found to bind to the enhancer core homology of polyoma virus in a manner qualitatively indistinguishable from C/EBP (data not shown).

The most prominent form of the protein that survived purification from frozen rat liver migrated on denaturating electrophoresis gels with an apparent molecular weight of 14 kD. Roughly 3 μg of this polypeptide (200 pmoles) were purified and subjected to gas-phase amino acid sequence analysis (performed by Ned Siegel at the Monsanto Corporation). These efforts yielded a partial amino acid sequence 17 residues in length. Cycle recoveries ranged from 149 to 39 pmoles, and averaged 70 pmoles (not including three residues that quantitate poorly: histidine, arginine, and threonine), indicating that the sequence reflected the major species that was recovered from purification. The sequence of these 17 amino acids was:

$$\text{H}_2\text{N-AGPHPDLRTGGGGGGGA-COOH}$$

Computer searches through several national data bases failed to reveal substantial similarity between this sequence and any known protein. Thus, other than revealing an unusually long stretch of glycines, the partial amino acid sequence of the 14-kD fragment of C/EBP was relatively uninformative.

Using the partial amino acid sequence of C/EBP as a guide, we synthesized a corresponding peptide (see Materials and methods). The synthetic peptide extended 13

residues from the amino-terminal alanine (through the fourth glycine), and terminated with a cysteine at the carboxyl terminus of the peptide. The peptide was purified by reverse-phase HPLC and coupled via the carboxy-terminal cysteine to keyhole limpet hemocyanin (KLH) using the heterobifunctional cross-linking reagent, N-(γ-buteryloxymalimido)-succinamide (GMBS), in a manner similar to that described by Green et al. (1982). Conjugated protein was injected into rabbits, leading to the production of an antiserum specific to the synthetic 14-amino-acid peptide. Evidence of specificity of the antiserum was first detected by radioimmunoassays comparing uncoupled KLH to conjugated material (data not shown). Two subsequent tests confirmed the specificity and usefulness of the anti-peptide serum. First, the serum was found to be capable of immuno-precipitating C/EBP binding activity. Second, it was capable of specific binding to C/EBP on western blots.

To test for specific immunoprecipitation, crude nuclear extracts were incubated with protein A–Sepharose beads containing bound IgG from either preimmune or immune serum. Protein bound to the beads was recovered by centrifugation and eluted at low pH (see Materials and methods). As Figure 1 shows, anti-peptide serum led to the selective precipitation of a DNA-binding activity that was capable of producing DNase I footprints on the two enhancer core motifs of the hepatatis B virus enhancer. The selectivity of immunoprecipitation was supported by two criteria. First, preimmune serum prepared from the same rabbit failed to precipitate any DNA-binding activity detectable in our assay. Second, the anti-peptide serum led to the precipitation of enhancer core binding activity to the exclusion of other DNA-binding activities that occur in crude rat liver nuclear extracts.

The anti-peptide serum was also observed to be capable of specific interaction on western blots with both the 14-kD polypeptide that was used for partial amino acid sequencing and the 20-kD form that was originally purified from fresh rat liver (data not shown). Western blot analyses that will be presented in the Results indicated that both the 20- and 14-kD species represent degradation products of a larger polypeptide. Given three indications of specificity (radioimmunoassay, immunoprecipitation, and western blotting), we proceeded to use the anti-peptide serum as a probe to screen a bacteriophage λgt11 library prepared from rat liver cDNA.

Isolation of cDNA and genomic C/EBP clones

Polyadenylated RNA isolated from adult rat liver was used to synthesize cDNA for insertion into λgt11. Recombinant phage were packaged in vitro resulting in a primary library of 2×10^6 phage. Phage were plated at a density of roughly 5×10^3 plaques/15-cm petri dish and screened by conventional procedures (Young and Davis 1983). One antibody-positive phage was recovered from a screen of roughly 2×10^5 plaques.

Four criteria indicated that this phage, termed L20, contains a cDNA copy of the mRNA encoding C/EBP.
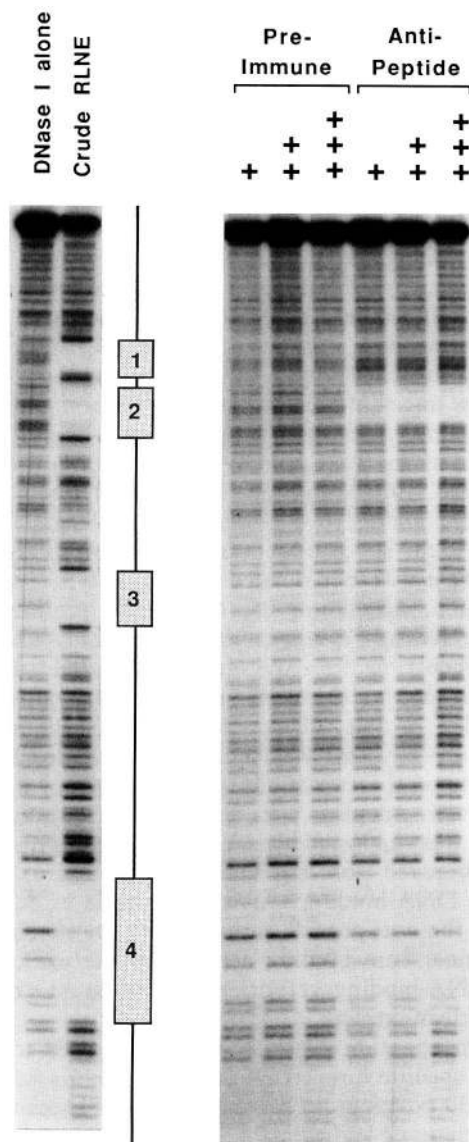
Figure 1. Immunoprecipitation of C/EBP from crude rat liver nuclear extract with anti-peptide serum. (*Left panel*) DNase I digestion patterns generated on a fragment of the HBV enhancer either in the absence or presence of rat liver nuclear extract (RLNE). The schematic diagram shows the locations of four DNase I footprints created by the crude RLNE. Sites 2 and 4 are known to represent binding sites for purified C/EBP (see Fig. 3). (*Right panel*) DNase I digestion patterns generated after exposure of the HBV enhancer DNA fragment to protein fractions that were immunoprecipitated from crude RLNE (see Materials and methods). Note that proteins eluted from the anti-peptide lead to selective protection of the probe DNA in regions corresponding to C/EBP binding sites (footprint regions 2 and 4). Region 2 corresponds to base pair coordinates 1032–1048 and region 4 to coordinates 1184–1209 of the HBV enhancer (Jameel and Siddiqui 1986; Shaul and Ben-Levy 1987).

The first two criteria are documented by data included in Figure 2. Infection of bacterial cells with L20 resulted

in the synthesis of a β-galactosidase fusion protein that migrated more slowly in denaturing electrophoresis gels than the form specified by the parental phage (λgt11). By comparing the mobilities of the two proteins to molecular weight standards, we estimated the product specified by L20 to exceed the mass of its gt11 counterpart by between 15- and 20-kD (Fig. 2a). The initial observation supporting the relatedness of the L20 fusion protein to C/EBP came from western blot analyses. Unlike the β-galactosidase polypeptide specified by gt11, the L20 fusion protein reacted with the anti-peptide serum. As shown in Figure 2b, the ingredient necessary for positive reactivity on western blots was absent in preimmune serum. Furthermore, the reactivity of immune serum was blocked by the synthetic peptide, but not by purified KLH carrier protein.

The second indication of relatedness between the L20 product and C/EBP came from a southwestern blot (Miskimins et al. 1985). Proteins from infected bacterial cells were again sized by SDS gel electrophoresis and transfered to nitrocellulose. In this case, however, filters were exposed to synthetic, ^{32}P-labeled DNA corresponding to a known C/EBP binding site (see Materials and methods). Native C/EBP can be detected by this method so long as 50 or more pmoles are added per gel lane (S. McKnight and B. Graves, unpubl.). To control for nonspecific interactions between protein and DNA, a parallel nitrocellulose filter was probed with ^{32}P-labeled salmon sperm DNA. The results of these binding assays, presented in Figure 2c, showed that the fusion polypeptide specified by L20 binds the synthetic DNA ligand with substantial preference over nonspecific DNA. nonspecific DNA.

A more rigorous test of sequence-specific DNA binding, DNase I footprinting, was performed as a third test of the authenticity of L20. Crude extracts prepared from bacterial cells infected with L20 were denatured in 5 M urea and chromatographed over DEAE cellulose to remove nucleic acids. Flowthrough fractions were dialyzed to remove the urea and tested for enhancer core binding activity by DNase I footprinting assays (see Materials and methods). As Figure 3 shows, extracts prepared from L20-infected bacterial cells contained an activity that bound to the enhancer core homologies of several viral enhancers. Side-by-side comparisons show no qualitative differences between the footprints produced by native C/EBP and that specified by the L20 bacteriophage.

Bacterial extracts were also examined for binding specificity to the CCAAT homology of the Moloney murine sarcoma virus (Mo-MSV) long terminal repeat (LTR). Earlier studies had shown that the avidity of C/EBP interaction with the CCAAT homology of the Mo-MSV LTR was increased when the CCAAT pentanucleotide carried a C → G transversion at the first residue of the consensus, and decreased when it carried an A → C transversion at its third residue (Graves et al. 1986). As shown in Figure 4, the product of L20 bound to the CCAAT homology of the Mo-MSV LTR in a manner indistinguishable from native C/EBP. The qualitative di-
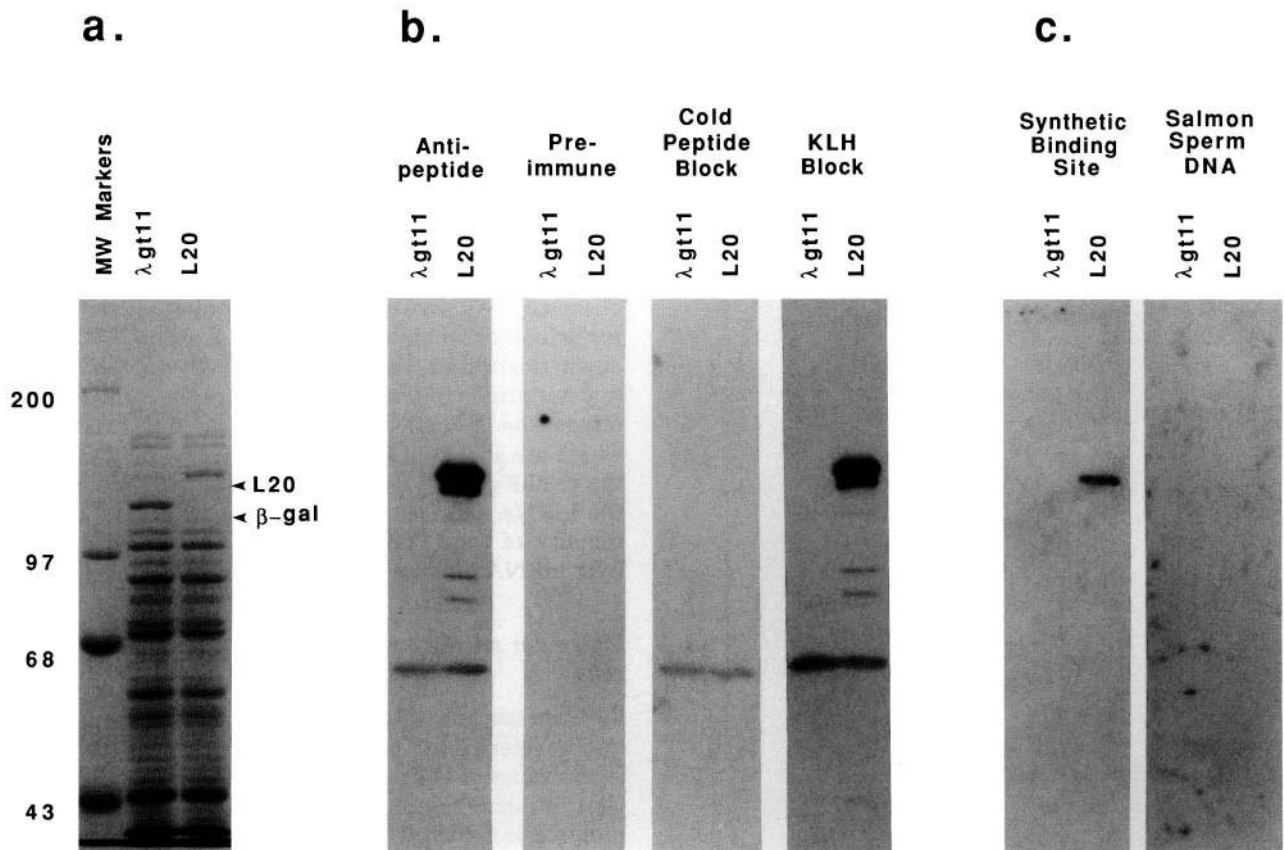
Figure 2. Western and southwestern blot analysis of the β-galactosidase fusion protein specified by bacteriophage λ recombinant L20. (*a*) Coomassie blue staining patterns of SDS gel-separated proteins that were extracted from *E. coli* cells infected with either bacteriophage λgt11 or the L20 recombinant. (*b*) Results of western blots of the same proteins that are visualized in the left panel. Note that the anti-peptide serum detects the fusion protein specified by the L20 bacteriophage (along with presumed protease degradation products), and that this interaction is blocked by excess synthetic peptide, but not by excess carrier protein (KLH). (*c*) Results of southwestern blots of the same proteins that are visualized in the left panel. Note that the synthetic C/EBP-specific probe binds specifically to the β-galactosidase fusion protein specified by the L20 bacteriophage (see Materials and methods for details).

mensions of the DNase I footprint on the 'wild-type' CCAAT sequence could not be distinguished from the footprint produced by the native protein. Furthermore, binding avidity of the bacterially produced protein was affected by the two point mutations in a manner concordant with the native protein.

The final criterion that established relatedness between L20 and C/EBP came as a result of DNA sequence analysis of the cDNA insert housed within the recombinant bacteriophage. A 1.2-kb insert was excised from purified L20 DNA by digestion with *EcoRI* and subcloned into a plasmid vector. A library of deletions was created to facilitate DNA sequencing by the chain-terminating method (see Materials and methods). The DNA sequence of the insert derived from L20 is presented in Figure 5a. Two aspects of this sequence were consistent with properties expected of a fusion protein between β-galactosidase and C/EBP. First, the translational reading frame that started at the fusion point between β-galactosidase and insert progressed for 146 codons before encountering a UGA translation termination codon. This

coding sequence, taking amino acid content into account, should add 16 kD of mass to β-galactosidase. Such calculations are consistent with the experimentally determined increase in mass of the L20 fusion protein relative to the form of β-galactosidase specified by gt11 (see Fig. 2). Second, this same translational reading frame exhibited a segment of 17 amino acids identical to the sequence obtained from the 14-kD fragment of C/EBP (Fig. 5b).

Using the insert of L20 as a hybridization probe, we retrieved four additional clones from our rat liver cDNA library. The insert of one of these clones, termed L39, was 2.3 kb in length and terminated with a poly(A) segment. DNA sequence analysis of L39 added 152 codons to the ORF of the L20 cDNA clone. Northern blot analysis (presented below) showed the C/EBP mRNA to be 2.7 kb in length. Since we were unable to retrieve any cDNA clones of this size, we used the L20 insert to screen a genomic library of rat liver DNA (Materials and methods). These efforts led to the isolation of a bacteriophage carrying a 9-kb *EcoRI* restriction fragment corre-
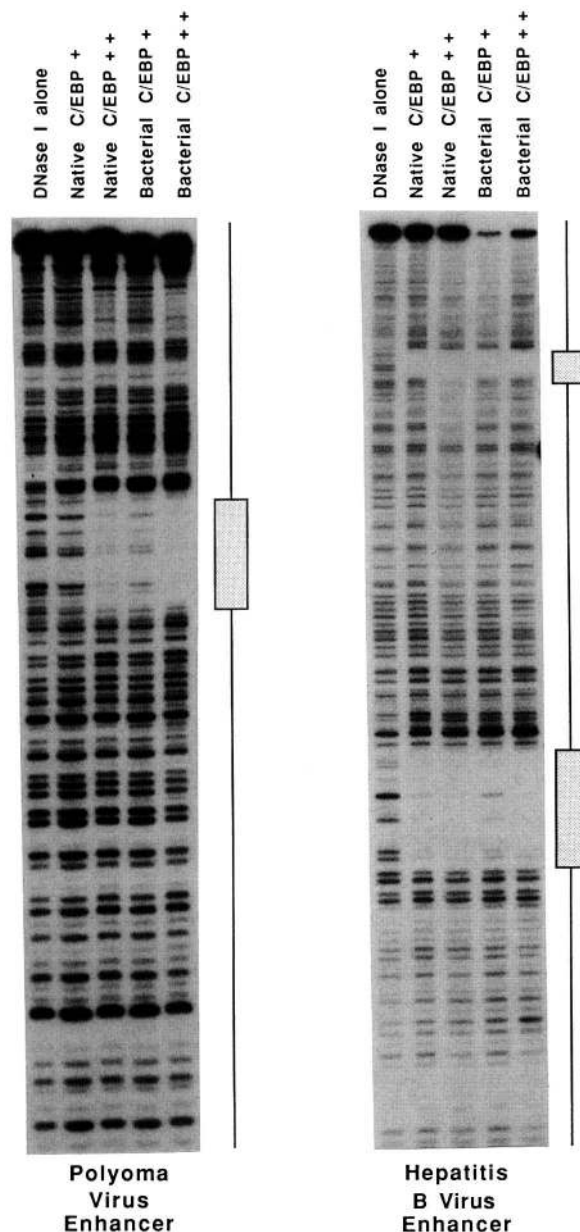
Figure 3. Comparison of footprinting patterns established on two viral enhancers by purified rat liver C/EBP and L20-specified fusion protein. DNA probes from the polyoma virus enhancer and HBV enhancer were exposed to either purified C/EBP (native C/EBP) or *E. coli* extracts prepared from cells infected with the L20 bacteriophage (bacterial C/EBP). Note that the two protein sources produced qualitatively indistinguishable patterns of DNase I protection on both enhancer DNA fragments. The DNA sequences underlying each of the observed DNase I footprint are related to the enhancer core sequence (see Johnson et al. 1987).

A canonical TATAAA homology was observed in the genomic DNA sequence 162 bp upstream from the first methionine codon of the C/EBP ORF. If transcripts of the gene were initiated 25 bp downstream from this TATAAA homology, and appended with 100 adenine residues at the polyadenylation site observed on the L39 cDNA clone, they would correspond to the experimentally measured size of the C/EBP mRNA (see subsequent section of Results). To test the possibility that C/EBP transcripts actually initiated in a manner directed by the observed TATAAA homology, we synthesized an oligonucleotide complementary to a region 44–70 bp downstream from the putative mRNA cap site and carried out primer extension assays on rat liver mRNA (Materials and methods). The results of this experiment, which are shown in Figure 5c, confirmed the location of the C/EBP mRNA cap site. Since the distance between the mRNA cap site and polyadenylation site matched the size of the C/EBP mRNA, we predict that the rat liver mRNA product of the C/EBP gene is unspliced.

Molecular properties of the C/EBP protein, mRNA, and gene

We have used the antibody and cloned DNA as reagents to characterize the C/EBP polypeptide further, and initiate studies of the gene and mRNA that encode it. In an effort to determine the size of undegraded C/EBP protein, we isolated rat liver nuclei in the presence of protease inhibitors, and rapidly extracted nuclear proteins by treatment with SDS gel sample buffer. Proteins were electrophoresed, transferred to nitrocellulose, and probed sequentially with rabbit anti-peptide serum and ^{125}I-labeled protein A. These procedures identified two polypeptide species, one that migrated with an apparent molecular mass of 42 kD and a second less abundant species that migrated an apparent mass of 30 kD (Fig. 6a). The 20- and 14-kD species of C/EBP were not observed under these conditions of nuclear isolation and sample preparation. Two observations indicated that the smaller forms of C/EBP represented degradation products of the 42-kD species. First, they only occur when precautions are not taken to prevent proteolysis. Second, DNA sequence analysis of cDNA and genomic clones has revealed an ORF of 359 codons (see Fig. 5a). Taking amino acid composition into account, this 359-codon ORF should specify the synthesis of a 38-kD polypeptide.

Independent of the aforementioned observations, there are three reasons to predict that the 42-kD species represents a less degraded form of C/EBP than was characterized in earlier studies. First, extracts containing predominantly the 42-kD species and no detectable 20- and 14-kD proteins exhibit the canonical C/EBP footprints. These extracts retained activity after being heated to 90°C, and the 42-kD protein remained soluble. Second, it exhibited chromatographic properties indistinguishable from the 20-kD form of C/EBP. Third, its reactivity with the anti-peptide serum was fully blocked by unlabeled peptide (data not included).

sponding to the gene encoding C/EBP. DNA sequence analysis of this genomic clone further extended the L20/L39 ORF to a methionine codon 359 residues upstream of the putative translation termination codon of C/EBP.

The size of the C/EBP mRNA and reiteration frequency of the gene that encodes it were studied using a radioactive probe derived from the insert of the L20 bacteriophage. Figure 6, b and c, show the results of northern and Southern blots, respectively. Poly(A)⁺ rat liver RNA contained a single, 2.7-kb mRNA species complementary to the C/EBP cDNA probe. Likewise, Southern blot analysis of rat liver DNA digested with either *Bgl*II or *Eco*RI revealed only a single DNA fragment complementary to the cDNA probe. Both northern and Southern hybridizations were performed and washed under stringent conditions. As such, these experiments did not address the issue of related genes or mRNAs.

Amino acid sequence of the C/EBP DNA binding domain

Analyses of proteolyzed fragments of C/EBP have localized its DNA-binding domain to a 14-kD fragment. Since the amino-terminus of this fragment had been identified by direct amino acid sequencing, we were able to establish the location of the 14-kD fragment on the linear map of C/EBP generated by DNA sequence analysis (see Fig. 5). The amino-terminal residue of the 14-kD fragment, an alanine, is located 106 residues upstream from the putative translation termination codon. Taking amino acid composition into account, these 106 codons specify a mass of 12 kD, indicating that the 14-kD fragment probably extends to the carboxyl ter-
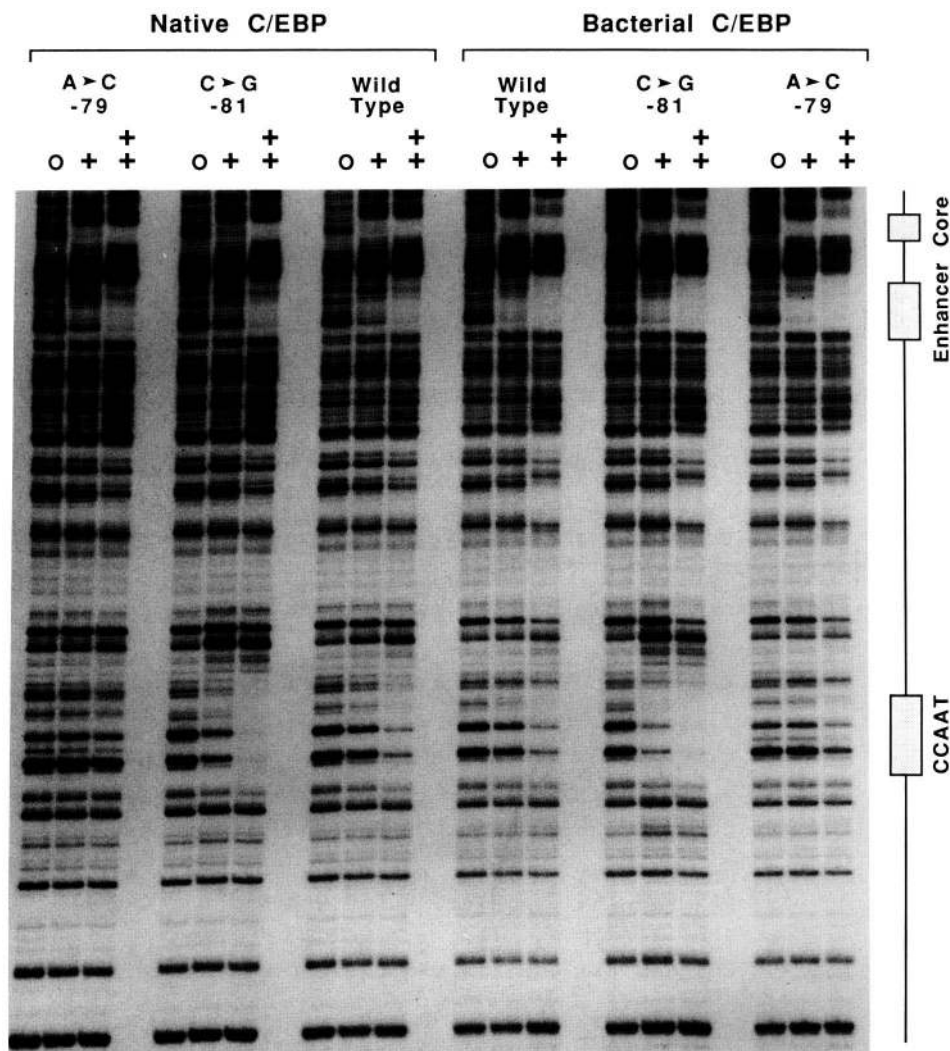
Figure 4. Comparison of DNase I footprint patterns established on the Mo-MSV LTR by native and bacterially-synthesized C/EBP. DNA fragments of the 'wild-type' Mo-MSV LTR and variants bearing single point mutations at two locations within the CCAAT homology were labeled, purified, and subjected to DNase I footprint analysis using either rat liver C/EBP (native C/EBP) or protein produced in *E. coli*. One of the point mutations in the CCAAT homology, a C → G transversion at the first residue of the pentanucleotide, was known to elevate the binding affinity of C/EBP (Graves et al. 1986). The other point mutation, an A → C transversion at the third residue of the pentanucleotide, was known the interfere with C/EBP affinity. These observations, reproduced in the left three panels, are recapitulated by footprint assays of bacterially synthesized C/EBP, which are shown in the right three panels.

```
                                          +1
          TATAAAAGCG GTACCCCGTC AGCGGGCCTG GCCATTCGCG ACCCAAAGCT GCGCGGGGTC    30

          AGCGGGCGAG TTGGGGTACT GGGTGGGCGG CGGCGACAGC GGCGCCACGC GCAGGCTGGA    90

                                                            M   E   S   A   D    5
          GGCCGCCGAG GCTCGCCATG CCGGGAGAAC TCTAACTCCC CC  ATG GAG TCG GCC GAC   147

           F   Y   E   A   E   P   R   P   P   M   S   S   H   L   Q   D   P     22
          TTC TAC GAG GCG GAG CCG CGG CCC CCG ATG AGC AGC CAC CTC CAG AGC CCC   198

           P   H   A   P   S   N   A   R   L   W   L   S   P   G   R   G   P     39
          CCG CAC GCG CCC AGC AAC GCC CGC CTT TGG CTT TCC CCG GGG CGC GGG CCC   249

           R   A   A   P   S   P   T   C   R   P   G   A   A   G   R   I   C     56
          CGC GCC GCC CCC AGC CCC ACC TGC CGC CCC GGA GCC GCT GGG CGC ATC TGC   300

           E   H   E   T   S   I   D   I   S   A   Y   I   D   P   A   A   F     73
          GAG CAC GAG ACG TCT ATA GAC ATC AGC GCC TAC ATC GAC CCG GCC GCC TTC   351

           N   D   E   F   L   A   D   L   F   Q   H   S   R   Q   Q   E   K     90
          AAC GAC GAG TTC CTG GCC GAC CTC TTC CAG CAC AGC CGG CAG CAG GAG AAG   402

           A   K   A   A   A   G   P   A   G   G   G   G   D   F   D   Y   P    107
          GCC AAG GCG GCG GCG GGC CCC GCG GGT GGC GGC GGT GAC TTT GAC TAC CCG   453

           G   A   P   A   G   P   G   G   A   V   M   S   A   G   A   H   G    124
          GGC GCC CCG GCG GGC CCC GGC GGT GCG GTC ATG TCC GCG GGG GCG CAC GGA   504

           P   P   P   G   Y   G   C   A   A   A   G   Y   L   D   G   R   L    141
          CCC CCT CCC GGC TAC GGC TGT GCG GCG GCC GGC TAC CTG GAC GGC AGG CTG   555

           E   P   L   Y   E   R   V   G   A   P   A   L   R   P   L   V   I    158
          GAG CCC CTG TAC GAG CGC GTC GGG GCG CCC GCG CTG CGG CCG CTG GTG ATC   606

           K   Q   E   P   R   E   E   D   E   A   K   Q   L   A   L   A   G    175
          AAG CAG GAG CCC CGC GAG GAG GAC GAG GCG AAG CAG CTG GCG CTG GCC GGC   657

           L   F   P   Y   Q   P   P   P   P   P   P   P   P   H   P   H   A    192
          CTC TTC CCC TAT CAG CCC CCG CCG CCG CCG CCG CCA CCG CAC CCG CAC GCG   708

           S   P   A   H   L   A   A   P   H   L   Q   F   Q   I   A   H   C    209
          TCT CCC GCG CAC TTG GCC GCC CCT CAC TTG CAG TTC CAG ATC GCA CAC TGC   759
          β-gal ──────────┤  ├────── L20 ──────────►
           G   Q   T   T   M   H   L   Q   P   G   H   P   T   P   P   P   T    226
          GGC CAG ACC ACC ATG CAC CTG CAG CCT GGC CAC CCT ACG CCG CCG CCG ACG   810

           P   V   P   S   P   H   P   A   P   A   M   G   A   A   G   L   P    243
          CCC GTG CCC AGC CCT CAT CCC GCG CCT GCA ATG GGT GCT GCG GGC CTG CCG   861

           G   P   G   G   S   L   K   G   L   A   G   P   H   P   D   L   R    260
          GGC CCC GGG GGC TCG CTC AAG GGC TTG GCT GGT CCG CAC CCC GAC CTC CGC   912

           T   G   G   G   G   G   G   G   A   G   A   G   K   A   K   K   S    277
          ACC GGC GGC GGC GGC GGC GGC GGG GCC GGC GCG GGC AAG GCC AAG AAG TCG   963

           V   D   K   N   S   N   E   Y   R   V   R   R   E   R   N   N   I    294
          GTG GAT AAG AAC AGC AAC GAG TAC CGG GTA CGG CGG GAA CGC AAC AAC ATC  1014

           A   V   R   K   S   R   D   K   A   K   Q   R   N   V   E   T   Q    311
          GCG GTG CGC AAG AGC CGA GAT AAA GCC AAA CAG CGC AAC GTG GAG ACG CAG  1065

           Q   K   V   L   E   L   T   S   D   N   D   R   L   R   K   R   V    328
          CAG AAG GTG TTG GAG TTG ACC AGT GAC AAT GAC CGC CTG CGC AAG CGG GTG  1116

           E   Q   L   S   R   E   L   D   T   L   R   G   I   F   R   Q   L    345
          GAA CAG CTG AGC CGT GAA CTG GAC ACG CTG CGG GGT ATC TTC CGC CAG CTG  1167

           P   E   S   S   L   V   K   A   M   G   N   C   A   ***                359
          CCT GAG AGC TCC TTG GTC AAG GCC ATG GGC AAC TGC GCG TGA GGCGCGCGGCT  1220

          GCGGGACCGC CTTGAGCCGG CCCCCTGGCT GGGGACCCGG AGGATGGTTT AGGTCGCTGG  1280
```

Figure 5. (*See following page for legend.*)

minus of C/EBP.

Having established a rough localization of the DNA-binding domain of C/EBP, we examined its amino acid sequence with the hope of finding motifs commonly observed in DNA-binding proteins. The minimal DNA-binding domain contains only a single cysteine (one residue from the putative carboxy terminus) and a single histidine (four residues downstream from the protease cleavage point that generated the 14-kD fragment). The paucity of these amino acids is likely to rule out utilization of the zinc finger motif (reviewed recently by Evans and Hollenberg 1988). Searches for sequence similarity to the helix-turn-helix motif, using the conventions of Pabo and Sauer (1984), were equally futile.

```
ATCTCTAGGC NCCGACCGTG CAAGCCAGAC TAGAGATTCT GGTGTCGCCT GAAGCCTGGC   1340

CTGCTCCGCG TGTCCCCTCC CTTCCTCTGA GCCGGACTCG GTGCGTCTAA GATGAGAGGG   1400

AGTCAGGCCG TGGTGGTTTC TCCTTTAGAC CGAGAGACTT CTCGCGGAGC TGAGCTGGGA   1460

GCCCGGCAGT TTTAGTATTA AGGAAGTAAC CTTGTGCCTT GGATACTCAA AACTCGCTCC   1520

TTTTCCTACC GAGTAGGGGG AGCAAACATG TGCCTTGATA TTTTATTTGG AGGATTCCTG   1580

CTTCCTCCAG AGCTGCAGCA GGCCCCCATG AGAAAAATGA AGGGTGCAGG CCCAGGGCAG   1640

GAGGAAGATT CAGGAAGCTG AGGTCCCAGC AGTCCCTGAG CTGCCCCTCC GTCCCTGTCC   1700

TTAGAGGGGA GGGACTTAGG GAGTTGGGGA TTTGAGTCTT AAAATTGTTC CCCACCCCTA   1760

GTTACATGAA TGAGGTCTAG CCTGCTTGCT CTATCTACAT TCCTCCCACC CAGCTTACAA   1820

CACAGTTCCT TGGTGAGTTC GTGGAGAAGG GGGCAACACC CCCAGTCAGA CCAGAAGGCT   1880

GAGTTGTGAA CTGGCCATGT GGTTAAGCAA GTGGTAGAAG ACAGGGACCT TAGTTTCTGG   1940

TCTTGGGGTG GAGGATGAAG ACACTGGGAC CATCCGCCTT GTGTGTACTG TATGTCGCCA   2000

GCCGCTGTTG CTGAAGGAAC TTGAAGCATA ATCGATCCAT CCCAGAGGGA CTGGAGTTAT   2060

GACAAGCTTC CCAAATATTT TGCTTTATCA TCTGATATCA ACACTTGTAT CTGGTCTCAG   2120

TGTCCCAGCG GTGCCTTGTG CAATGTCAGT TGCGCGTCTA TCTAAACCAC CATTTTATTT   2180

GGTCTTTTGT TTTGTTTTGG TTTTGCTCTG ATTCTTGCCA AAATGAGACT CTTCACGAAC   2240

GGCCAGGGGA AGGGGCTGAG TGAGGCTCTC ATTCTTTTTG GTTTAGGGAT GTTTGGTTTT   2300

TTTGTTTTGT TTTGTTTTGT TTTGTTTTCT TTCGCCTCCA GAGGACCAAT GAAATGAAGT   2360

GGGCTTCTCC CTCTCCCCTA GTTTTCCAAG GGGTTAGTTA GTTGTGGGTC TTAGCTTCCT   2420

TCTCCGGGCT AGACTTATAC TGGCTTCCCC ACCTACCCAC CCCCATCCCC AAAGGCCCTG   2480

GCTCTGGGTC TGGAAAGAAG GCCACCTCCA GCCAATTTGT ACACAACCCT GCCACTGGAA   2540

GCAGGGGTGG ACTGCTTCCT TCTCTTCTTT TTTTGGGGGG AGGGGACACA AAGTTTCATG   2600

CTAGATCTCC TATGTATTAT ATCTATAATA TAAACATATC AAACTCAAAA A            2651
```

Figure 5. DNA sequence of the gene ecoding C/EBP. (*Left panel*) Nucleotide sequence of the rat gene that encodes C/EBP. The sequence was derived from both cDNA and genomic clones (see Materials and methods). The nucleotide sequence starts at a TATAAA homology located 162 bp upstream from the putative translation initiation codon of C/EBP. The location of the TATAAA homology allowed prediction of the mRNA cap site at the guanine residue that is labeled + 1. The legitimacy of this prediction was tested using a synthetic oligonucleotide complementary to residues + 44 to + 70 of the C/EBP mRNA. The primer was annealed to rat liver mRNA and exposed to reverse transcriptase to facilitate primer extension mapping of the C/EBP mRNA cap site. The results of this assay, which is shown in (next page) *c*, confirm the location of the mRNA cap. (*b*) Autoradiogram of a sequencing gel that contains the DNA sequence of the C/EBP gene in a region corresponding to the amino terminus of the 14-kD peptide that yielded partial amino-terminal amino acid sequence. The location of this sequence, which corresponds to amino acid residues 253–268 of the C/EBP ORF, is highlighted on the panel showing the DNA sequence of the gene. The fusion point of the L20 cDNA clone to the β-galactosidase gene of bacteriophage λgt11 occurs in frame with the C/EBP ORF at threonine residue 212. DNA sequence analysis of four additional cDNA clones indicate that the C/EBP ORF is followed by a long 3′ untranslated region (1442 residues in length). The cDNA containing the largest extent of the C/EBP ORF (termed L39) terminated with 5 adenine residues, which were preceded by a canonical polyadenylation sequence (AATATAAA).
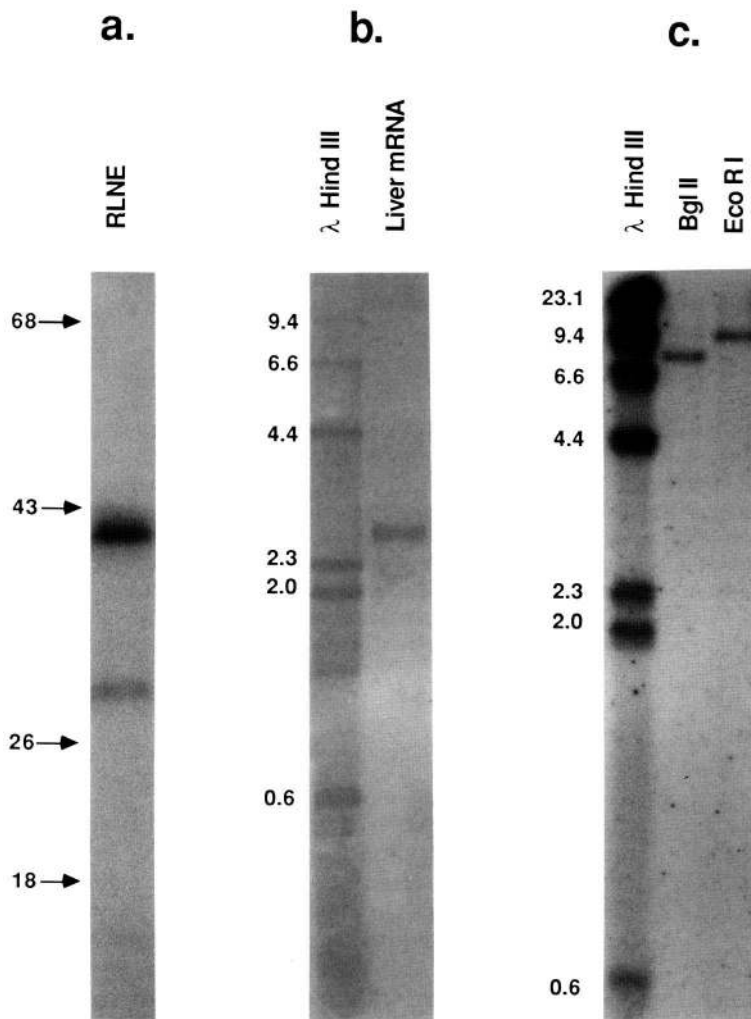
Although the DNA-binding domain of C/EBP was likely to be distinct from either of the two existing prototypes, it did exhibit the high density of basic amino acids expected of a DNA-binding protein. If we assume that the single cysteine is reduced, and ignore the potential contributions of posttranslational modifications (acetylation and phosphorylation), then the 14-kD fragment would exhibit an isoelectric point of 11, and have a net charge of +10 at pH 7.0. The spatial distribution of basic amino acids in the region of C/EBP that encompassed its DNA binding domain was markedly nonrandom (Fig. 7). Five clusters of basic amino acids occurred within a 55-amino-acid span. Nine amino acids separated the first cluster (KAKK) from the second cluster (RVRR), seven separated the second from the third cluster (RKSR), one residue separated the third cluster from the fourth cluster (KAKQR), and 16 separated the third from the fourth cluster (RLRKR). We speculate that these clusters of basic amino acids play a role in the interaction between C/EBP and DNA.

Although we had not noticed any similarity between the amino acid sequence of C/EBP and either of the prototypical DNA binding motifs (helix-turn-helix or zinc fingers), a computer search revealed substantial similarity between C/EBP and the products of the c-*myc* and v-*fos* oncogenes. Both of these transforming proteins are known to reside in the nucleus, and evidence has been presented that they are capable of binding DNA (review recently by Eisenman and Thompson 1986). In the case of v-*fos*, the region of strongest similarity coincided with the second cluster of basic amino acids within the DNA-binding domain of C/EBP (Fig. 7). In the case of

Figure 5. (*See preceding page for legend.*)

Figure 6. Western, northern, and southern blot analysis of the C/EBP polypeptide, mRNA, and gene. The autoradiogram presented in *a* shows a western blot of rat liver nuclear protein that, to prevent proteolytic degradation, was exposed to SDS gel sample buffer immediately after extraction from fresh tissue. The predominant cross-reacting species exhibits an apparent molecular weight of 42 kD. The autoradiogram presented in *b* shows a northern blot of rat liver mRNA. RNA was extracted from adult rat liver after solubilization in guanidine hydrochloride and purified from contaminating protein by organic extraction. Polyadenylated RNA was recovered by affinity chromatography on oligo-(dT)-cellulose and electrophoresed on a denaturing agarose gel. The hybridization probe used to detect the 2.7-kb C/EBP mRNA was the cDNA insert of the L20 bacteriophage recombinant. The autoradiogram presented in *c* shows a Southern blot of rat liver DNA probed with the cDNA insert of L20. Restriction with both *Eco*RI and *Bgl*II leads to the generation of a single DNA fragment that hybridizes to the L20 probe.

c-*myc*, the region of strongest similarity coincided with the fifth cluster of basic amino acids. Both regions of similarity were compact (11 identities out of 15 for c-*myc*, and 9 out of 15 for c-*fos*), and uninterrupted by either the addition or deletion of amino acids. The only amino acid sequence similarity that we observed to be shared between the three proteins is a periodic repeat of leucines that spans the fifth cluster of basic amino acids of C/EBP. Speculations on the role of this leucine repeat motif have been presented elsewhere (W.H. Landschulz, P.F. Johnson, and S.L. McKnight 1988).

Discussion

The results presented in this report allow several new conclusions to be drawn relative to the nature of C/EBP. First, it can now be stated with certainty that this one protein is capable of interacting with two *cis*-regulatory elements that share no overt similarity at the DNA sequence level (the CCAAT homology common to many promoters and the core homology common to many en-

hancers). This conclusion is reinforced most substantially by the observation that both CCAAT and enhancer core binding activities are specified by protein synthesized in *E. coli* from a single cloned gene. Now that we have access to the gene encoding C/EBP, it should be possible to use molecular genetic methods to dissect functional domains of the protein. Using such methods, we hope to test whether identical, overlapping, or distinct regions of C/EBP are required for sequence-specific interaction with its disparate recognition sites on DNA.

A second set of observations that have emerged from these studies concerns the sizes of the intact C/EBP polypeptide and the mRNA that encodes it. When precautions were taken to prevent proteolysis, the major immunoreactive species migrated on denaturing electrophoresis gels with a relative molecular mass of 42 kD. Apparently, the 20-kD species that was purified in an earlier study (Johnson et al. 1987) represents a partially degraded form of the protein. We do not know whether the 42-kD form observed on SDS gels represents the size
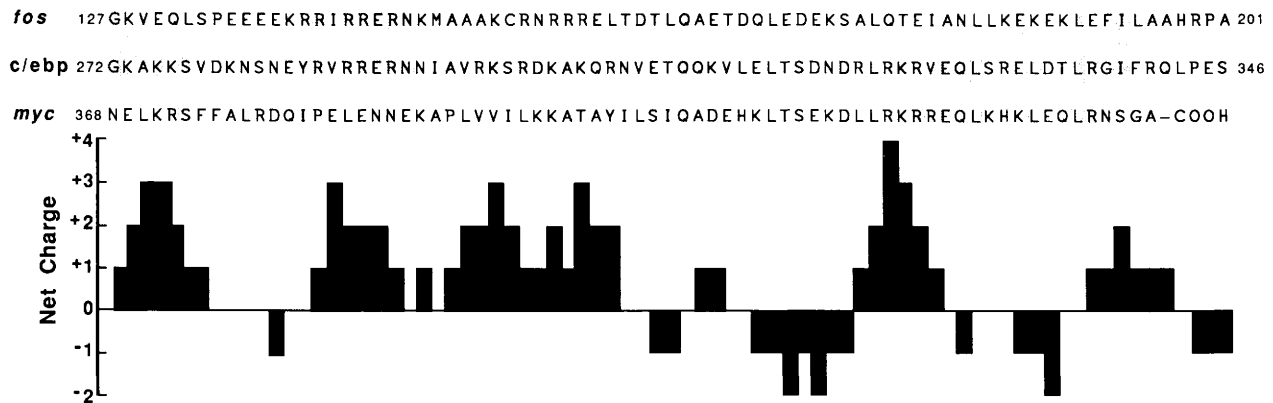
Figure 7. Amino acid charge distribution of the DNA-binding domain of C/EBP and comparisons to the amino acid sequences of the *myc* and *fos* oncoproteins. [*Top*] Amino acid sequences of C/EBP in a region encompassing its DNA-binding domain. Regions of sequence similarity between C/EBP, v-*fos*, and c-*myc* were determined by a computer search of the National Biomedical Research Fund protein data base. The most notable relatedness between C/EBP and v-*fos* is localized to a highly basic region spanning a 5-amino-acid identity (RRERN). The most notable relatedness between C/EBP and c-*myc* is also localized to a highly basic region spanning a 4-amino-acid identity (LRKR). The only substantial relatedness between the three proteins is a heptad repeat of leucines spanning the region that is most similar between C/EBP and c-*myc*. [*Bottom*] Net charge density in windows of 5 amino acids across the DNA-binding domain of C/EBP. This analysis identifies five peaks of basic charge density (containing a net charge of +3 within a 5-amino-acid window). Note that two such peaks correspond to regions of sequence similarity between C/EBP and the two oncoproteins.

of the intact polypeptide. However, since the size of the mRNA is 2.7 kb and since roughly 1400 bases of the mRNA occur in 3' untranslated sequences (see Figs. 5a and 6), the intact protein is not likely to be considerably larger than 42 kD.

The third and perhaps most surprising observation that has emerged from this report concerns the sequence relatedness between C/EBP and the products of the c-*myc* and v-*fos* oncogenes. There are several reasons to believe that the observed similarities are of functional significance. First, they occur within the 14-kD segment of C/EBP that includes its DNA-binding domain. Second, it is known that carboxy-terminal truncations of the c-*myc* protein that extend into C/EBP-related amino acid sequences fully abrogate transforming activity of the protein (Stone et al. 1987). Indeed, Stone et al. (1987) reported that a 2-amino-acid insertion in c-*myc* within its region of similarity to C/EBP completely eliminates its transforming activity. A previous study noted amino acid sequence similarity between c-*myc*, v-*fos*, and a newly discovered oncogene termed v-*jun* (Vogt et al. 1987). Various lines of evidence have indicated that the product of the v-*jun* oncogene is a sequence-specific DNA-binding protein (Struhl 1987; Bohmann et al. 1987). The coincidence of related amino acid sequences in three nuclear transforming proteins and the DNA-binding domain of C/EBP may reflect the underpinnings of a common structural motif that facilitates DNA binding.

What we have failed to provide, thus far, is any compelling evidence of the functional involvement of C/EBP in the regulation of gene expression. Now that we have access to a recombinant DNA clone of the gene en-coding C/EBP, it may be possible to study its regulatory role in vivo. For example, we should be able to test its effects on putative target genes by transient transfection into cultured mammalian cells. Alternatively, mutated forms of the C/EBP protein may be engineered in an attempt to create a *trans*-dominant form of the protein that would interfere with normal C/EBP activity.

Although lacking concrete support, our proposition that C/EBP will play a positive role in transcriptional regulation is consistent with its affinity for *cis*-regulatory elements commonly associated with genes that encode mRNA. Our own identification of C/EBP binding sites has been limited to viral enhancers and promoters (Graves et al. 1986; Johnson et al. 1987; Figs. 1 and 3 of this study). However, published work from three other laboratories has indicated that C/EBP also plays a role in cellular gene expression. Elegant in vitro transcription studies of the albumin promoter have revealed the involvement of a heat-stable protein indistinguishable from C/EBP (Lichsteiner et al. 1987). The site within the albumin promoter that binds this factor is critical to albumin gene expression both in vivo (Heard et al. 1987) and in vitro (Gorski et al. 1986; Lichsteiner et al. 1987). Purified C/EBP, whether obtained from rat liver or *E. coli*, binds avidly to the albumin promoter in precisely the same region identified in the aforementioned studies (P. Johnson, S. Cereghini, and M. Yaniv, unpubl.). Along these same lines, Darnell and colleagues have found that *cis*-regulatory elements associated with two other liver-specific genes, α1-antitrypsin and transthyretin, bind both the heat-stable form of C/EBP and the bacterially derived product of our recombinant clone (Costa et al. 1988).

In closing, we return to the enigma of enhancer–promoter communication. If, as we suspect, C/EBP does act as a positive regulator of transcription, then what role does the protein play once it binds to DNA? We do not notice any particularly acidic regions that might correspond to the transcriptional activating domains of GAL4, GCN4, and VP16 (Hope and Struhl 1986; Ma and Ptashne 1987; Treizenberg et al. 1988). Since the degree of acidity of an 'acid blob' appears to correlate with its activating potency (Gill and Ptashne 1987; Triezenberg et al. 1988), C/EBP, unless substantially modified by phosphorylation, might constitute a rather weak activator if it acts by this convention.

What other ways might C/EBP carry out its presumed activating role? Computer searches have failed to reveal any substantial amino acid sequence similarity between the amino-terminal segment of C/EBP, which is not essential for DNA binding, and amino acid sequences that exist in the National Biomedical Research Fund protein data base. Such observations are only useful in ruling out obvious potential functions (e.g., protein kinase activity). A plausible mechanism for C/EBP function that has not yet been eliminated is as a mediator of direct interaction between enhancers and promoters. Since C/EBP appears capable of binding to cis-regulatory sequences found in both enhancers and promoters, it is possible that it facilitates the direct physical juxtaposition of enhancers and promoters. Since enhancers are believed to consist of binding sites for a number of different proteins, we offer the speculation that certain proteins (e.g., C/EBP) facilitate communication with a promoter, and others (e.g., those appended with an acidic activating surface) function more directly to induce gene expression. This idea requires, at a minimum, that C/EBP be capable of multimerizing. Since we are now able to purify large quantities of bacterially synthesized C/EBP, it should be possible to test this prediction directly.

Materials and methods

Preparation of C/EBP for amino-terminal sequencing

Crude rat liver nuclear extract (RLNE) was prepared from ~200 frozen rat livers (Sprague-Dawley retired breeders, Pel-freeze) as described by Graves et al. (1986). C/EBP was purified from the RLNE as described by Johnson et al. (1987) through the Mono-S cation-exchange column stage. Eluant fractions containing the appropriate C/EBP DNA-binding activity were electrophoresed on a 10–20% polyacrylamide gradient SDS gel. The use of frozen rat liver led to the generation of several proteolyzed fragments with apparent molecular weights between 10 and 20 kD. C/EBP candidate polypeptides were excised and isolated by electroelution as described previously (Johnson et al. 1987). An aliquot of each preparation was tested for enhancer core binding activity using a DNA probe derived from polyoma virus. A clearly resolved 14-kD polypeptide, which bound avidly to the polyoma virus probe, was successfully sequenced by Dr. Ned Siegel (Monsanto Corp.). The protein was applied to a precycled Polybrene filter disk and subjected to automated Edman degradation (Hunkapillar et al. 1983) using an Applied Biosystems 470A Protein Sequencer. The PTH amino acid derivatives were analyzed online using an Applied Biosystems 120A PTH analyzer.

Preparation of anti-C/EBP serum

A 14-amino-acid peptide with the sequence (NH_2)-AGPHPDLRTGGGGC-(CO_2H) was sythesized by Dr. Clarke Riley (Howard Hughes Medical Institute, Johns Hopkins University School of Medicine) using t-BOC chemistry on an Applied Biosystems 430A Peptide Synthesizer. The peptide was purified by reverse-phase HPLC and coupled via its free sulfhydryl group to keyhole limpet hemocyanin (KLH). Ten milligrams of KLH was dissolved in 0.5 ml of 10 mM sodium phosphate buffer (pH 8.0). Free sulfhydryl groups on KLH were inactivated by treatment with 5 μg of N-ethylmalaimide (NEM) for 1 hr, leaving free amino group available to react with the hetero-bifunctional cross-linker γ-maleimidobutyric acid N-hydroxysuccinimide ester (GMBS). After adjusting the pH with an equal volume of 10 mM phosphate buffer (pH 7.0), 2 mg GMBS (dissolved in 5 μg of dimethylformamide) was added slowly to the stirring KLH solution. Excess, unreacted reagents were removed 30 min later by gel filtration over G-50 in 10 mM phosphate buffer (pH 6.0). The activated KLH was mixed with 10 mg of the synthetic peptide in 0.5 ml of 10 mM phosphate (pH 6.0). Coupling was allowed to proceed for 2 hr at room temperature. The coupling efficiency was monitored using fluorescamine-labeled peptide (Udenfriend et al. 1972). Approximately 60% of tracer peptide was found associated with KLH, indicating that each KLH molecule was coupled to approximately 10 peptides. NEM, GMBS, fluorescamine, and KLH were purchased from Sigma Chemical Co.

The peptide-coupled KLH was used to immunize rabbits. Seven-week-old, New Zealand White rabbits were injected initially with 2 mg of the antigen, and boosted at 2-week intervals with 0.2 mg of the antigen. The serum used in this report was obtained 3 months after the initial injection.

Specificity of the sera was tested by immunoprecipitation of C/EBP activity in a manner modified from the methods of Schneider et al. (1982). To minimize proteolytic degradation of C/EBP, crude rat liver nuclear extracts were prepared by the methods of Gorski et al. (1986) using freshly dissected liver and supplementing the nuclear isolation and resuspension buffers with protease inhibitors (0.5 mM PMSF, 2.5 mM benzamidine, 10 μg/ml aprotinin, 20 μg/ml trypsin inhibitor, and 1 μg/ml each Chymostatin, leupeptin, antipain, and pepstatin). Two hundred microliters of packed Protein A–Sepharose beads (Pharmacia) were incubated with 750 μl of either preimmune or immune serum. The beads were then washed twice with PBS, twice with immunowash buffer [0.5 M NaCl, 50 mM Tris (pH 8.2), 1 mM EDTA, 0.5% NP-40], and three times with PBS. Crude RLNE was diluted 1 : 1 with HGEDK50 [10 mM HEPES (pH 7.9), 10% glycerol, 1 mM EDTA, 1 mM DTT, 50 mM KCl], and 200 μl of this dilution was mixed with the beads and incubated at 4°C. After 4 hr, the beads were isolated by centrifugation, and washed twice with immunowash buffer, and twice with 0.5 M NaCl. Antibody-antigen complexes were solubilized by two sequential 40 μl washes of 0.18 M glycine (pH 2.8), 150 mM NaCl, and 10% dioxane. After neutralization with 20 μl 0.5 M Na_2HPO_4, the solution was tested for DNase I protection activity as described earlier (Graves et al. 1986).

Isolation of recombinant clones encoding C/EBP

Poly(A)$^+$ RNA was isolated from rat liver (retired breeder, Sprague-Dawley rats, Charles River Laboratories, Inc). Five micrograms of poly(A)$^+$ RNA was used to synthesize cDNA with

a cDNA synthesis kit (Amersham). Two micrograms of bacteriophage λgt11 DNA cut with EcoRI (Promega Biotek), were ligated to 400 ng of cDNA. The concatenated DNA was packaged in vitro using a commercial extract (Stratagene). A cDNA library containing 2×10^6 total plaques was obtained. The library was not amplified before use. Lambda λgt11 expression screening was carried out according to published procedures (Young and Davis 1983) using horseradish peroxidase conjugates (Vector Labs) to visualize positive clones. A single positive clone, termed L20, was identified from a screen of roughly 200,000 plaques.

Preparation of bacterial extracts containing C/EBP

E. coli extracts were prepared in three ways. First, E. coli extracts were prepared for electrophoretic sizing and subsequent western and southwestern blotting. These extracts were made from cells infected with either λgt11 or L20. An overnight culture of E. coli strain Y1089 was diluted and grown in log phase until it reached a density of 0.4 O.D. units (at 600 nm). Cells were collected by centrifugation, resuspended in 0.2 volumes of fresh medium, and infected with either λgt11 or L20 at a multiplicity of infection of 5. After adsorption for 10 min at room temperature, twice the original volume of medium, prewarmed to 42°C, was added and the culture was grown in a shaker for 10 min at 42°C. Isopropylthiogalatoside (IPTG) was then added to 10 mM and the cultures were transferred to a 37°C shaker incubator. Fusion protein synthesis was allowed to proceed for 1 hr. Cells were harvested by centrifugation, lysed in SDS sample buffer, and heated for 5 min at 100°C. Cell debris was removed by centrifugation and solubilized protein was loaded onto SDS electrophoresis gels (Laemmli 1970).

Second, phage-infected cell extracts were prepared for DNase I footprint analysis. Again, infected and induced cells were harvested by centrifugation. In this case, however, the cells were resuspended in cold Tris-buffered saline [TBS, 10 mM Tris (pH 7.5), 150 mM NaCl], that was supplemented with 0.05% Triton-X-100, 1 mM benzamidine, 1 mM EDTA, and 2.5 M urea. Resuspended cells were lysed by sonication. Insoluble debris and nucleic acids were removed by increasing the urea concentration to 5 M, and running the sonicated material over DEAE–Cellulose (Whatman). Flowthrough fractions were dialyzed against buffer free of urea, TGEDK100, and used in footprinting reactions as described previously (Graves et al. 1986).

Third, extracts from E. coli engineered to express C/EBP free of β-galactosidase were prepared for footprint analysis. A cDNA insert encoding the carboxy-terminal 30-kD of C/EBP was cloned downstream of the bacteriophage T7 phi-10 gene promoter as described by Studier and Moffett (1986). In this case, cells were grown to mid-log phase, induced with IPTG, and harvested. Extracts were prepared for footprinting as described above.

Western and southwestern analysis

Bacterial and rat liver nuclear extracts were electrophoresed on SDS gels and transferred to nitrocellulose. Western blot analysis was performed as described by Towbin et al. (1979) using [125I]Protein A (Amersham) to visualize immunoreactive material. Nitrocellulose filters were probed for sequence-specific DNA binding activity by the method of Miskimins et al. (1985). A synthetic C/EBP binding site was prepared by annealing the two oligonucleotides shown below:

```
5'-AATTCAATTGGGCAATCAGG - 3 '
   3 ' - GTTAACCCGTTAGTCCTTAA-5'
```
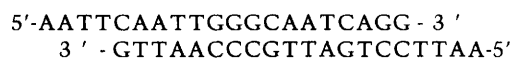
After phosphorylating the two oligonucleotides with T4 polynucleotide kinase and unlabeled ATP, they were annealed and concatenated to an average length of 200 nucleotides using T4 DNA ligase (Vinson et al. 1988, this issue). This concatenated DNA was radiolabled by nick-translation (Maniatis et al. 1983). Sonicated, salmon sperm DNA was used as a probe for nonspecific binding. After proteins were transferred onto nitrocellulose, the filter was blocked with 5% (wt/vol) nonfat dry milk in binding buffer [10 mM HEPES (pH 7.9), 50 mM KCl, 1 mM EDTA, 1 mM DTT] for 1 hr at 4°C. The filters were transferred to binding buffer supplemented with 0.25% dry milk and 100,000 cpm of ^{32}P-labeled DNA per milliliter, and incubated for 1 hr at 4°C. Filters were washed in three 100-ml changes of binding buffer, 0.25% milk at 4°C (10 min per wash). After washing, filters were blotted on 3MM paper (Whatman) and exposed to X-ray film.

Sequencing of C/EBP clones

The insert from the positive clone, L20, was subcloned into the Bluescript vector (Stratagene), and a series of deletions starting from the 3' end of the cDNA were made. Deletion mutants were sequenced using the chain-termination method and Sequenase reagents (U.S. Biochemical Corp.) Regions of the DNA sequence that were not clearly resolved were subjected to chemical sequencing (Maxam and Gilbert 1980).

The sequence of the genomic clone of C/EBP showed a TATAAA homology 132 bp upstream from the putative translation initiation codon of C/EBP. The authenticity of the TATAAA homology was tested using a primer extension assay. A synthetic primer complementary to the mRNA 69–94 residues downstream from the TATAAA homology was synthesized. The primer was labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase, hybridized to poly(A)$^+$ rat liver RNA, and exposed to reverse transcriptase using the conditions of McKnight et al. (1981). Primer extension products were sized on an 8% sequencing gel and visualized by autoradiography.

Acknowledgments

References

Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27: 299–308.

Banerji, J., L. Olson, and W. Schaffner. 1983. A lymphocyte-specific enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33: 729–740.

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 transcription factor AP-1. *Science* **238**: 1386–1392.

Costa, R.H., D.R. Grayson, K.G. Xanthopoulos, and J.E. Darnell, Jr. 1988. A liver-specific DNA binding protein recognizes multiple nucleotide sites in regulatory regions of transthyretin, alpha 1-antitrypsin, albumin, and SV40 genes. *Proc. Natl. Acad. Sci.* (in press).

Dandanell, G. and K. Hammer. 1985. Two operator sites separated by 599 base pairs are required for *deoR* repression of the deo operon of *Escherichia coli*. *EMBO J.* **4**: 3333–3338.

Dunn, T.M., S. Hahn, S. Ogden, and R.F. Schleif. 1984. An operator at −280 base pairs that is required for repression of *araBAD* operon promoter: Addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci.* **81**: 5017–5020.

Eisenman, R.N. and C.B. Thompson. 1986. Oncogenes with potential nuclear function: *myc, myb,* and *fos. Cancer Surv.* **5**: 309–329.

Evans, R.M. and S.M. Hollenberg. 1988. Zinc fingers: Gilt by association. *Cell* **52**: 1–3.

Fromm, M. and P. Berg. 1983. Simian virus 40 early- and late-region promoter fuctions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Mol. Cell. Biol.* **3**: 991–999.

Gill, G. and M. Ptashne. 1987. Mutants of GAL4 protein in an activation function. *Cell* **51**: 121–126.

Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa. 1983. A tissue-specific transcriptional enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**: 717–728.

Gorski, K., M. Carneiro, and U. Schibler. 1986. Tissue-specific in vitro transcription from the mouse albumin promoter. *Cell* **47**: 767–776.

Graves, B.J., P.F. Johnson, and S.L. McKnight. 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. *Cell* **44**: 565–576.

Green, N., H. Alexander, A. Olson, S. Alexander, T.M. Shinnick, J.G. Sutcliffe, and R.A. Lerner. 1982. Immunogenic structure of the influenza virus hemagglutinin. *Cell* **28**: 477–487.

Hager, D.A. and R.R. Burgess. 1980. Elution of proteins from sodium dodecyl sulfate, and renaturation of enzymatic activity: Results with sigma subunit of *E. coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.* **109**: 76–86.

Heard, J.M., P. Herbomel, M.O. Ott, A. Mottura-Rollier, M. Weiss, and M. Yaniv. 1987. Determinants of rat albumin promoter tissue specificity analyzed by an improved transient expression system. *Mol. Cell. Biol.* **7**: 2425–2434.

Herr, W. and J. Clarke. 1986. The SV40 enhancer is composed of multiple functional elements that can compensate for one another. *Cell* **45**: 461–470.

Hochschild, A.H. and M. Patashne. 1986. Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell* **44**: 681–687.

Hope, I. and K. Struhl. 1986. Functional dissection of a eukaryotic transcription activator protein, GCN 4 of yeast. *Cell* **46**: 885–894.

Hunkapillar, M.W., R.M. Hewick, W.J. Dreyer, and L.E. Hood. 1983. High-sensitivity sequencing with a gas-phase sequenator. *Methods Enzymol.* **91**: 399–413.

Jameel, S. and A. Siddiqui. 1986. The human hepatitis B virus enhancer requires *trans*-acting cellular factor(s) for activity. *Mol. Cell. Biol.* **6**: 710–715.

Johnson, P.F., W.H. Landschulz, B.J. Graves, and S.L. McKnight.

1987. Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. *Genes Dev.* **1**: 133–146.

Kraemer, H., M. Niemoeller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann, and B. Mueller-Hill. 1987. *lac* repressor forms loops with linear DNA carrying two suitably spaced *lac* operators. *EMBO J.* **6**: 1481–1491.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Lee, W., A. Haslinger, M. Karin, and R. Tjian. 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature* **325**: 368–372.

Lichtsteiner, S., J. Wuarin, and U. Schibler. 1987. The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. *Cell* **51**: 963–973.

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

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manal.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Maxam, A. and W. Gilbert. 1980. Sequencing end-labeled DNA with base specific chemical cleavages. *Methods Enzymol.* **65**: 499–560.

McKnight, S.L., E.R. Gavis, R. Kingsbury, and R. Axel. 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: Identification of an upstream control region. *Cell* **25**: 385–398.

Miskimins, W.K., M.P. Roberts, A. McClelland, and F.H. Ruddle. 1985. Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor. *Proc. Natl. Acad. Sci.* **82**: 6741–6744.

Mitchell, P.J., C. Wang, and R. Tjian. 1987. Positive and negative regulation of transcription in vitro: Enhancer-binding protein AP-2 is inhibited by SV40 T-antigen. *Cell* **50**: 847–861.

Moreau, P., B. Wasylyk, R. Everett, M.P. Gaug, and P. Chambon. 1981. The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* **9**: 6047–6068.

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

Piette, J., M.-H. Kryszke, and M. Yaniv. 1985. Specific interaction of cellular factors with the B enhancer of polyoma virus. *EMBO J.* **4**: 2675–2685.

Prywes, R. and R.G. Roeder. 1986. Inducible binding of a factor to the c-*fos* enhancer. *Cell* **47**: 777–784.

Queen, C. and D. Baltimore. 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. *Cell* **33**: 741–748.

Reitzer, L.J. and B. Magasanik. 1986. Transcription of *glnA* in *E. coli* is stimulated by activator bound to sites far from the promoter. *Cell* **45**: 785–792.

Schirm, S., J. Jiricny, and W. Schaffner. 1987. The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificty. *Genes Dev.* **1**: 65–74.

Schneider, C., R.A. Newman, D.R. Sutherland, U. Asser, and M.F. Greaves. 1982. A one-step purification of membrane proteins using a high efficiency immunomatrix. *J. Biol. Chem.* **257**: 10766–10769.

Shaul, Y. and R. Ben-Levy. 1987. Multiple nuclear proteins in liver cells are bound to hepatitis B virus enhancer element and its upstream sequences. *EMBO J.* **6**:1913–1920.

Speck, N.A. and D. Baltimore. 1987. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol. Cell. Biol.* **7:** 1101–1110.

Stone, J., T. de Lange, G. Ramsay, E. Jakobovits, J.M. Bishop, H. Varmus, and W. Lee. 1987. Definition of regions in human c-*myc* that are involved in transformation and nuclear localization. *J. Mol. Cell. Biol.* **7:** 1697–1709.

Struhl, K. 1987. The DNA binding domains of the *jun* oncoprotein and the yeast GCN4 transcriptional activator proteins are functionally homologous. *Cell* **50:** 841–846.

Studier, W.F. and B.A. Moffett. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189:** 113–130.

Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci.* **76:** 4350–4354.

Treisman, R. 1986. Identification of a protein-binding site that mediates transcriptional response of the c-*fos* gene to serum factors. *Cell* **46:** 567–574.

Treisman, R., M.R. Green, and T. Maniatis. 1983. *cis* and *trans* activation of globin gene transcription in transient assays. *Proc. Natl. Acad. Sci.* **80:** 7428–7432.

Triezenberg, S.J., R.C. Kingsbury, and S.L. McKnight. 1988. Functional disection of VP16, the *trans* activator of Herpes simplex virus immediate early gene expression. *Genes Dev.* **2:** 718–729.

Undenfriend, S., S. Stein, P. Boehlen, W. Dairman, W. Leimgruber, and M. Weigele. 1972. Flourescamine: A regeant for assay of aminoacids, peptides, proteins, and primary amines in the picomole range. *Science* **178:** 871–872.

Vogt, P.K., T.J. Bos, and R.F. Doolittle. 1987. Homology between DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxy-terminal region of a protein coded for by the oncogene *jun*. *Proc. Natl. Acad. Sci.* **84:** 3316–3319.

Walker, M.D., T. Edlund, A.M. Boulet, and W.J. Rutter. 1983. Cell-specific expression controlled by the 5'-flanking region of insulin and chymotrypsin genes. *Nature* **306:** 557–561.

Wildeman, A.G., M. Zenke, C. Schatz, M. Wintzerith, T. Grundstrom, H. Mattes, K. Takahashi, and P. Chambon. 1986. Specific protein binding to the simian virus 40 enhancer in vitro. *Mol. Cell. Biol.* **6:** 2098–2105.

Young, R.A. and R.W. Davis. 1983. Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci.* **80:** 1194–1198.

Zenke, M., T. Grundstrom, H. Mattes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon. 1986. Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* **5:** 387–397.

Isolation of a recombinant copy of the gene encoding C/EBP.

W H Landschulz, P F Johnson, E Y Adashi, et al.

References	This article cites 53 articles, 17 of which can be accessed free at: **http://genesdev.cshlp.org/content/2/7/786.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.**