Rous Sarcoma Virus Enhancer Factor I Is a Ubiquitous CCAAT Transcription Factor Highly Related to CBF and NF-Y*

(Received for publication, May 31, 1990)

Mary Faber$ and Linda Sealy±

From the Department of Molecular Physiology and Biophysics and $Department of Cell Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

We have characterized enhancer factor I (EFI), a trans-acting factor present in avian nuclear extracts which binds to the Rous sarcoma virus long terminal repeat enhancer and promoter. Through deletion and point mutagenesis, we show that EFI is a member of the CCAAT family of transcription factors. Although the CCAAT motif is essential for protein-DNA recognition, EFI shows surprising latitude in the nucleotide sequences flanking the CCAAT motif to which it will bind with high affinity. EFI will cross-bind to the binding sites of a number of previously described CCAAT factors, including CBF, NF-Y, CP2, CP1, CTF/NF-1, and c/EBP, with a range of affinities that is at most 10-fold lower than the high affinity binding site for EFI in the Rous sarcoma virus long terminal repeat. We present evidence, however, that EFI is probably identical or very closely related to CBF and NF-Y. This is based on the fact that EFI in avian nuclear extracts binds with equal or 2-fold greater affinity to the binding sites of NF-Y and CBF, despite less than 50% homology (outside the CCAAT motif) between the EFI, NF-Y, and CBF recognition sequences. Moreover, radiolabeled EFI, NF-Y, or CBF DNAs give rise to identical gel retardation patterns in extracts from a variety of different cell types. EFI, CBF, and NF-Y appear to fractionate identically upon ion exchange chromatography, separating into two heterologous components (A and B) which must be recombined to recover substantial DNA binding activity. Molecular weight estimates for the two heterologous components of EFI, CBF, and a Y-box binding protein (Celada, A., and Maki, R. A. (1989) Mol. Cell. Biol. 9, 3097±3100) are very similar. EFI DNA binding activity has recently been shown to be induced by serum and the oncogene v-src (Dutta, A., Stoeckle, M. Y., and Hanafusa, H. (1989) Genes & Dev. 4, 243±254). The close relationship or identity between EFI, CBF, and NF-Y, thus has important implications regarding the mechanisms by which serum or the oncogene v-src may affect changes in gene expression.

The regulation of transcription initiation in eukaryotic cells is achieved through a complex interplay of multiple cis-acting DNA sequences and trans-acting protein factors. The regulatory DNA sequences which govern the expression of a particular gene have been broadly classified into two groups: promoter proximal elements and enhancers (Dyan and Tjian, 1985; Mitchell and Tjian, 1989; Dynan, 1989). As their name implies, promoter proximal elements are located near the start site of transcription whereas enhancers function independently of position, orientation, and distance with respect to the site of transcription initiation (Serfling et al., 1985; Puschke, 1986; Dynan and Tjian, 1989; Dynan, 1988).

Numerous transacting protein factors which bind in a sequence-specific manner to the promoter proximal elements or enhancers of a number of different viral and cellular genes have been identified (for reviews see Maniatis et al., 1987; Jones et al., 1988; Mitchell and Tjian, 1989), and some of these transcription factors have been characterized in great detail (Landschulz et al., 1988b; Courey and Tjian, 1988; Clerc et al., 1988; Hollenberg and Evans, 1988; Mermod et al., 1989; Mitchell and Tjian, 1989 and references therein). However, the molecular mechanism(s) by which these protein factors act to regulate the process of transcription initiation is not yet understood.

We have utilized the long terminal repeat (LTR) sequences of the avian retrovirus Rous sarcoma virus (RSV) as a model system for studying the molecular events involved in the regulation of transcription initiation. The enhancer and promoter elements located in the U3 region of the RSV LTR (Luciw et al., 1983; Laimins et al., 1984; Cullen et al., 1985a; Weber and Schaffner, 1985), in conjunction with a second enhancer element located in the gag-encoding sequences of the retrovirus (Arrigo et al., 1987; Karnitz et al., 1987), provide for extremely high levels of expression of the viral genome (Gorman et al., 1982; Cullen et al., 1985b; Varmus and Swanson, 1985 and references therein). The strong transcriptional activating abilities of the RSV LTR are not confined to the virus' natural avian cell host. The RSV LTR enhancer and promoter are active in a wide variety of vertebrate cell types (Gorman et al., 1982), making these cis-acting DNA sequences an attractive system for investigating the basic molecular mechanisms involved in transcriptional activation in eukaryotic cells.

As a first step in delineating these mechanisms, we have been identifying and characterizing the trans-acting protein factors which bind in a sequence-specific manner to the RSV LTR enhancer and promoter. The RSV LTR enhancer has been functionally defined by deletion mutagenesis (Luciw et al., 1983; Laimins et al., 1984; Cullen et al., 1985a) and

1 The abbreviations used are: LTR, long terminal repeat; RSV, Rous sarcoma virus; bp, base pair(s); EFI, II, and III, enhancer factor I, II, and III; SDS, sodium dodecyl sulfate; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MHC, major histocompatibility complex; EGTA, [ethylenediaminetetraacetic acid].

* This work was supported in part by United States Public Health Service Grant GM39826 and March of Dimes Grant 1189 (to L. S.).
† Supported by National Research Service Award GM12250.
‡ A Scholar of the Leukemia Society of America.
observations, coupled with our characterization of EFI by respectively, upon SDS-gel electrophoretic analysis. These CPP-specific and CPl-specific binding sites (Chodosh et al., 1988) encompassing a region from -54 to -229 bp upstream from the start site of transcription. Within this region, we have so far described three factors, EFI, EFII, and EFiII, which recognize specific nucleotide sequences occurring at -121 to -146, -203 to -229, and -87 to -112, respectively (Sealy and Chalkley, 1987; Boulden and Sealy, 1990). We have recently reported (Boulden and Sealy, 1990) on a number of similarities between the EFI recognition site in the RSV LTR and the NF-Y binding site (Jones et al., 1987) or c/EBP transcription factors NF-Y (Darn et al., 1987) and CBF (Hata- nomochi et al., 1988), respectively, despite only 33% homology between the EFI recognition site in the RSV LTR and the NF-Y binding site outside the CCAAT motif. EFI binds to a CTF/NF-I binding site (Chodosh et al., 1988) with only 2- or 6-fold lower affinity, respectively, and to a CTF/NF-1 binding site (Jones et al., 1987) or c/EBP binding site (Graves et al., 1986) with 10-fold lower affinity. We demonstrate that radiolabeled oligonucleotides representing the EFI, CBF, and NF-Y binding sites give rise to identical mobility shift patterns using nuclear extracts from chick embryos or from the sources previously shown to contain CBF and NF-Y (namely rat hepatocyte and P388 mouse macrophage cells). Furthermore, we find that the EFI, CBF, and NF-Y binding activities fractionate in an identical manner upon ion-exchange chromatography, separating into two components (factor A and factor B) which must be recombined to recover a majority of the original DNA binding activity. Finally, EFI, CBF/-NF Y, and EFI, CBF/-NF Y in chick embryo extracts appear to have the same molecular weights, respectively, upon SDS-gel electrophoretic analysis. The observed differences between the purification ofEFI by cDNA cloning to be reported elsewhere (Ozer et al., 1990) lead us to suggest that EFI, CBF, and NF-Y are the same, or a very highly related, CCAAT transcription factor.

MATERIALS AND METHODS

Nuclear Extracts—Nuclear extracts were prepared from 14-day-old chick embryos (SPAFA8 Inc., Preston, CT) as described in Boulden and Sealy (1990) except that the protease inhibitor benzamidine was included in all buffers. A 0.1 M NaCl (buffer B) extract of chick embryo nuclei was used for all the binding assays presented here. Nuclear extracts were prepared from the rat hepatoma cell line H416E and the mouse macrophage cell line P388 as described in Sealy and Chalkley (1987), with the following modifications. After preparation of nuclei in buffer A (Sealy and Chalkley, 1987), the nuclei were resuspended directly in buffer B containing 0.5 M NaCl and extracted for 30 min on ice with gentle vortexing every 10 min. After centrifugation at 27,000 x g for 20 min, the supernatant (0.5 M nuclear extract) was dialyzed as previously described. A HeLa whole cell transcription extract was prepared as described by Shapiro et al. (1988). The H416E extract was a generous gift of Tony Ip (Columbia University) and Roger Chalkley (Vanderbilt University). HeLa cells were a generous gift of Robert Roeder's laboratory (The Rockefeller University) and were grown in Joklik's medium supplemented with 5% calf serum; P388 cells were obtained from the American Type Culture Collection and were maintained according to the accompanying instructions.

Radiolabeled and Competitor DNAs—Oligonucleotides were synthesized on an Applied Biosystems model 380A DNA synthesizer and purified as described in Boulden and Sealy (1990) prior to use. Complementary strands were annealed also as described in Boulden and Sealy (1990). Nucleotide sequences of the oligonucleotides used in these studies are presented in Table I, except for the 41-bp synthetic EFI DNA which is shown as follows.

5'-cCgggagAGAAAAAGCACCGTGCATGCCGATTGGTGAAGTAc-3'

The 8002 and L59 DNAs are linker insertions of this sequence as shown in Fig. 1. Double-stranded oligonucleotides were radiolabeled with [32P]ATP and polynucleotide kinase. Full length radiolabeled oligonucleotides were then purified by electrophoresis on a native 12% polyacrylamide gel, followed by electrophoresion and ethanol precipitation of the DNA. Nonradiolabeled competitor oligonucleotides were also purified by polyacrylamide gel electrophoresis, and the amount of DNA recovered was quantitated by measuring the absorption at 280-nm prior to use in gel retardation assays.

Gel Retardation Assays—Samples of nuclear or transcription extraction were mixed with 1 ng of 32P-labeled oligonucleotide and 575-775 ng of poly(dI)-(poly(dC) (Pharmacia IKR Rintechology Inc.) prepared as previously described (Sealy and Chalkley, 1987). Samples were incubated at room temperature for 30 min in a final volume of 7.6 mL containing 7.5 m3 Tris, pH 8.3, 2.1 m3 EDTA, 7 m3 0.2 M mercaptoethanol, 60 m3 NaCl, and 10% glycerol. Samples were analyzed on polyacrylamide gels as previously described (Sealy and Chalkley, 1987). Where appropriate, the protein-DNA complexes, located by autoradiography, were excised from the dried gel and radioactivity in the complex quantitated by liquid scintillation counting. Background counts were obtained by counting an equivalent section of the dried gel outside the sample lanes. The error in analyzing duplicate samples by this method was approximately 10%.

Ion Exchange Chromatography—Q-Sepharose (Pharmacia) chromatography was performed as described by Hata- nomochi et al. (1988) with minor modifications. A 0.1 M NaCl chick embryo nuclear extract which had been dialyzed to 10 mm HEPES, pH 8.0, 0.1 mm EDTA, 50 mm NaCl, 7 mm 0.2 M mercaptoethanol (buffer Q) containing 50% glycerol was diluted to buffer Q with 5% glycerol, or a fresh nuclear extract was prepared in buffer B containing 5% glycerol. Extract from approximately one embryo was mixed for 4 h at 4°C with 0.2 mL (quart volume) of the protein A/H- Sepharose immobilized in buffer Q containing 5% glycerol. Unbound material (referred to as flow-through, although chromatography was performed batchwise) was collected by low speed centrifugation, and the resin was washed in buffer Q except containing 0.1 M NaCl and 5% glycerol. The bound material was eluted for 45 min in buffer Q containing 1 M NaCl and 5% glycerol. Fractions were assayed on the same day before dialysis or after overnight dialysis to buffer Q containing 50% glycerol. Solutions for ion exchange chromatography contained 5 mg/mL leupeptin, 0.1 mm pepstatin, 1 mm phenylmethylsulfonyl fluoride, and 10 mm benzamidine added just before use.
TABLE I

Sequences of CCAAT box oligonucleotides

Nucleotide sequences of the oligonucleotides employed as binding sites for the different CCAAT factors are shown. Wild type (wt) EFI, high affinity EFI binding site at −121 to −146 in the RSV LTR (Sealy and Chalkley, 1987); CBF, CBF binding site at −72 to −97 in the mouse α2(I)-collagen promoter (Hatamochi et al., 1986); NF-Y, sequences −44 to −69 in the mouse MHC class II K, gene (Dorn et al., 1987); CP2, sequences −64 to −80 of the mouse MHC H-2Kk gene promoter (Chodosh et al., 1988); CP1, sequences −69 to −98 of the mouse α-globin gene promoter (Dorn et al., 1987; Cohen et al., 1986) (−71 to −94 of the mouse α-globin promoter are identical in sequence to −58 to −81 of the human α-globin promoter employed by Chodosh et al., 1988); CTF/NF-I, 28-bp CTF/NF-I binding site in the upstream region of the human ras-1 (Ha-ras) gene (Jones et al., 1987); c/EBP, −68 to −97 of the herpes simplex virus thymidine kinase kinase promoter (Graves et al., 1986). Sequences have been aligned by the inverted CCAAT motif which is enclosed by a box. Nucleotides in each CCAAT factor binding site homologous to the wild type EFI sequence have also been enclosed by boxes.

Sequences

<table>
<thead>
<tr>
<th>wt EFI</th>
<th>CBF</th>
<th>NF-Y</th>
<th>CP2</th>
<th>CP1</th>
<th>CTF/NF-I</th>
<th>c/EBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACGTGCATCGCGAGTTGGTGGAGAGTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCACGCTCTCCCATTTTGTTAAAGT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AATTGGTGGATTTTGTGCGTGCAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATGGTTGGTGTCGGGCGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCCTACATTTGCGCATGGGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGCGCTTGTATGGCAGCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A.

RESULTS

Insertion Mutagenesis of the EFI Binding Site—We have previously demonstrated that nuclear extracts of avian fibroblasts contain a protein factor, EFI, which specifically binds to and protects from DNase I attack an approximately 30-bp region (−121 to −146) in the RSV LTR enhancer in vitro (Sealy and Chalkley, 1987). The exact upstream boundary of protection was difficult to determine due to the lack of DNase I cutting of naked DNA in this region.) The nucleotide sequence of the LTR enhancer within the EFI-protected region is shown in Fig. 1A. To determine more precisely the nucleotide sequence actually recognized by EFI, we have generated several insertion mutants of the EFI binding site. In particular, synthetic 41-bp oligonucleotides representing either the wild type binding site or two linker substitution mutants (8902, LS9) shown in Fig. 1A were obtained for use in gel mobility shift assays. The 8902 mutant contains a 10-bp BamHI linker sequence (CCGGATCCGG) substituted for nucleotides −135 to −145 in the EFI binding site, whereas in the LS9 mutant, nucleotides −123 to −134 were replaced by the linker sequence. These oligonucleotides were radiolabeled and mixed with chick embryo nuclear extracts in the presence of increasing concentrations of poly(dI)-poly(dC) prior to electrophoretic analysis. As shown in Fig. 1B, both the wild type and 8902 mutant oligonucleotides formed a poly(dI)-poly(dC)-resistant complex with EFI, whereas the LS9 mutation completely eliminated EFI binding. This indicates that the nucleotides critical for EFI binding are contained within a 12-bp region asymmetrically disposed to the 3' side of the DNA protected by EFI from DNase I attack.

Point Mutagenesis of the EFI Binding Site—We noted that the DNA sequence altered in the LS9 mutation contained an inverted CCAAT motif at nucleotides −129 to −133. Since a number of previously characterized trans-acting factors have been shown to recognize the CCAAT motif (Graves et al., 1986; Cohen et al., 1986; Olesen et al., 1987; Jones et al., 1987; Dorn et al., 1987; Chodosh et al., 1988; Hatamochi et al., 1988; Gallinar et al., 1989), we carried out systematic point mutagenesis of this motif in the EFI binding site to determine its relevance for factor binding. Specifically, we obtained a series.

FIG. 1. Insertion mutagenesis of the EFI binding site. A, nucleotide sequence of the high affinity EFI binding site in the RSV LTR as defined by DNase I footprinting (Sealy and Chalkley, 1987). The inverted CCAAT motif is underlined. Open boxes below indicate the nucleotides replaced by the BamHI linker sequence CCGGATCCGG in the linker insertion mutants 8902 and LS9. B, radiolabeled 41-bp oligonucleotides representing the wild type high affinity EFI binding site at −121 to −146 (32P-WT), or the LS9 or 8902 mutants, were mixed with 2.5 μg of protein extracted from chick embryo nuclei by 100 mM NaCl, in the absence or presence of increasing amounts of poly(dI)-poly(dC) as shown. After electrophoresis of the samples on a 4% Tris-glycine-polyacrylamide gel, the gel was dried and exposed to x-ray film.
of synthetic 26-bp oligonucleotides containing either the wild type binding site (the wild type 26-bp sequence (−121 to −146) is equivalent to the 41-bp oligonucleotide containing the entire DNase I-protected region in binding EFI), or point mutations at each position of the inverted CCAAT box in turn, as well as selected point mutations outside of the inverted CCAAT box. These oligonucleotides were used in competition experiments to determine their ability to bind EFI in gel mobility assays. As shown in Fig. 2A and the summary graph in Fig. 2B, transversion mutations at three positions within the inverted CCAAT box (129 G to T, 131 T to G, and 133 A to C) eliminated EFI binding. Transversion mutations at the remaining two positions in the inverted CCAAT box (130 G to T and 132 T to G) significantly decreased EFI binding in vitro. Densitometric analysis of the autoradiograph shown in Fig. 2A indicated that the T to G transversion at position −132 reduced binding affinity approximately 60-fold, whereas the G to T transversion at position −130 had a more moderate effect, reducing EFI binding affinity approximately 15-fold. In contrast, two point mutations outside the inverted CCAAT motif (G to T transversions at −127 and −134) had no effect on EFI binding in vitro. These results identify EFI as a member of the CCAAT family of transcription factors.

Effect of Flanking Sequences on EFI Binding—Although the inverted CCAAT motif in the EFI binding site is clearly essential for factor binding, the CCAAT motif alone is not sufficient for specific protein-DNA recognition in vitro. This was demonstrated by analyzing the ability of EFI to bind in vitro to four inverted CCAAT boxes which occur elsewhere in the RSV genome. A computer-assisted search of the nucleotide sequence of the RSV genome (Schwartz et al., 1983) revealed 16 occurrences of inverted CCAAT boxes. For further analyses, we chose randomly two inverted CCAAT boxes which occur in the pol gene or gag gene, as well as two additional inverted CCAAT boxes which occur in the LTR. The additional inverted CCAAT boxes in the LTR are located in U5, and again in U3, at a site (−65 to −69) downstream of the EFI binding site. The nucleotide sequences of 26-bp oligonucleotides encompassing these four inverted CCAAT motifs, along with their specific location in the RSV genome, are given in Fig. 3A; competition assays with the inverted CCAAT box oligos are shown in Fig. 3B. EFI did not bind the inverted CCAAT box that occurs either in the pol gene or in U5. However, the inverted CCAAT box which occurs at −65 to −69 in U3 does serve as a binding site for EFI in vitro.

![Fig. 2. Point mutagenesis of the EFI binding site. A, 1 ng of 32P-labeled 41-bp EFI oligonucleotide was mixed with chick embryo nuclear extract as described under "Materials and Methods," in the absence or presence of a 2- to 100-fold molar excess of nonradiolabeled competitor DNAs. The competitor DNAs were 26-bp oligonucleotides representing the high affinity EFI binding site in the RSV LTR at −121 to −146 (wt) or point mutants at the positions indicated. After electrophoresis, the gel was dried and exposed to film. Autoradiography of that portion of the gel containing the EFI protein-DNA complexes is shown. B, to determine the binding affinity of the various EFI point mutant DNAs relative to the wild type binding site, the EFI protein-DNA complexes shown above were excised from the dried gel and radioactivity in the complexes was quantitated by liquid scintillation counting. The percentage of DNA bound in the presence of competitor DNA relative to that observed in the absence of competitor was calculated and plotted versus log(competitor DNA). The amount of point mutant competitor DNA required to reduce complex formation by 50% was calculated from this plot and divided by the amount of wild type competitor DNA required to reduce complex formation by 50% to obtain the values plotted in B.]

![Fig. 3. Effect of flanking sequences on EFI binding. A, nucleotide sequences of the competitor 26-bp oligonucleotides used in B are shown. Location of these nucleotide sequences in the RSV proviral genome (if appropriate) is also given; numbering is with respect to the start of transcription at +1. Nucleotides homologous to the wild type EFI binding site in the different oligonucleotides shown are shaded. B, a series of competition experiments were performed as described in the legend to Fig. 2, each employing a different competitor DNA as indicated. An autoradiograph of that portion of the gel containing the EFI protein-DNA complexes is shown.]
exhibiting a relative affinity approximately 5-fold lower than the upstream or wild type site at -120. This is consistent with other studies which have indicated that both CCAAT motifs in U3 play a role in promoter/enhancer function of the LTR (Dutta et al., 1990; Greuel et al., 1990). The inverted CCAAT motif in the gag gene is a relatively weak site for EFI binding, with an affinity approximately 20-fold below the highest affinity site in U3.

When examining the sequences flanking each of the three inverted CCAAT boxes (U3(-129), U3(-65), and 8902) which we have so far shown can serve as strong binding sites for EFI, we noted that the nucleotides to the left of the inverted CCAAT box can vary considerably. For example, only two of 11 nucleotides in the left-hand flanking sequences are conserved between the EFI binding sites at -129 and -65 in U3. Similarly, only three of 11 nucleotides to the left of the inverted CCAAT box match between the wild type and 8902 mutant EFI binding sites. The only nucleotide that each of these three EFI binding sites, U3(-129), U3(-65), and 8902, share in common to the left of the inverted CCAAT box is a G immediately flanking the A of the inverted CCAAT sequence. (However, this G is not required for EFI binding (point mutagenesis, Fig. 2), and it is not a site of close protein-DNA contact in methylation interference experiments (Table II).

The right-hand flanking sequences show greater conservation between the two EFI binding sites in U3. Four of eight nucleotides match, including the string GAA. Since both of the pol- and U5-inverted CCAAT box oligos lack these conserved nucleotides in their right-hand flanking sequences, we tested whether this difference was responsible for the lack of EFI recognition. A hybrid-inverted CCAAT box oligonucleotide was synthesized in which the right-hand flanking sequences were those found in the wild type U3(-129) EFI binding site and the left-hand flanking sequences were those of the pol-inverted CCAAT box. When tested in a competition assay, shown in Fig. 3, the hybrid wild type/pol-inverted CCAAT oligonucleotide bound EFI with only a 2-fold lower affinity than the complete wild type site. Thus, the conserved nucleotides in the righthand flanking sequences between the two EFI binding sites in U3 are not requisite for EFI recognition.

We have been unable in these, and further binding experiments described below, to identify any specific nucleotide in the sequences flanking the inverted CCAAT box which is required for EFI recognition. Although some constraints on flanking sequence must exist, since not all inverted CCAAT boxes are recognized by EFI, the requirements are presumably flexible. In other words, although additional protein-DNA contacts occurring in the sequences flanking the CCAAT box are required for EFI binding, no single contact is obligatory. It should be kept in mind that these are in vitro competition assays using 41- and 26-bp oligonucleotides. Conclusions drawn from these studies may not exactly mimic the binding of EFI in vivo, which may also depend on the context of the region, i.e. on what other transcription factors are bound to the RSV enhancer/promoter region, and possibly the regulatory state of each bound factor.

In the absence of a consensus recognition sequence (extending beyond the CCAAT box) for EFI binding in vitro, we were unable to predict the relationship between EFI and other CCAAT DNA binding proteins that have been previously characterized based upon a nucleotide comparison of their respective DNA binding sites. We therefore have obtained synthetic oligonucleotides representing different CCAAT factor binding sites to test directly whether EFI might be related to other CCAAT factors such as CTF/NF-1 (Jones et al., 1987), NF-Y (Dorn et al., 1987), CP1, CP2 (Chodosh et al., 1988), and others.

EFI Recognizes Other CCAAT Factor Binding Sites—The nucleotide sequences surrounding six different CCAAT boxes previously identified as specific, high affinity binding sites for the CCAAT binding factors CTF/NF-1 (Jones et al., 1987), CP1 or α-CP1 (Chodosh et al., 1988; Cohen et al., 1986), CP2 (Chodosh et al., 1988), NF-Y (Dorn et al., 1987), CBF (Hata-mochi et al., 1986), and c/EBP (Graves et al., 1986), respectively, are given in Table I. Synthetic oligonucleotides representing each of these CCAAT factor binding sites were employed in competitive experiments. Radiolabeled wild type U3(-129) 41-bp EFI DNA was mixed with increasing concentrations of the different CCAAT factor oligonucleotides over a range covering at least three orders of magnitude. After incubating these DNAs with an EFI-containing nuclear extract from chick embryos in the presence of poly(dI), poly(dC) and subsequent electrophoretic analysis, the binding isotherms present in Fig. 4 were constructed. The CTF/NF-1 and c/EBP binding sites exhibited a 10-fold lower affinity for EFI when compared to the highest affinity EFI site in U3. For comparison, the binding affinity of a point mutant, within the EFI inverted CCAAT box at -129, which has a binding affinity decreased by approximately two orders of magnitude, and the L9S mutant, which does not bind EFI, are presented on each graph for comparison. The high affinity binding site recognized by CP1 in the α-globin promoter showed an approximately 6-fold decrease in affinity for EFI, and the CP2-specific binding site found at position -72 of the H-2Kb major histocompatibility gene was only 2-fold lower in affinity. Interestingly, the NF-Y binding site showed an affinity identical to the wild type EFI site in U3, and the CBF binding site from the a2(l)-collagen promoter actually exhibited a 2-fold higher affinity than the wild type EFI site (129) in U3.

These results suggested that EFI may be identical or most closely related to the previously identified CCAAT factors CBF and NF-Y. It would be difficult to predict, a priori, that

| Table II |
| Comparison of methylation interference patterns of CCAAT binding factors |

Methylation interference patterns of: EFI from chick embryos on the high affinity (-129) binding site in the RSV LTR, CP1 from HeLa cells on the adenosine major late promoter, NF-Y from mouse B-lymphoma cells on the MHC class II E, gene promoter, and CP2 from HeLa cells on the γ-fibrinogen gene promoter. The CCAAT motif is enclosed by a box. Data are taken from Chodosh et al. (1988), Dorn et al. (1987), and Greuel et al. (1990). * strong interference; O, weak interference.
FIG. 4. EFI recognizes other CCAAT factor binding sites. A series of competition experiments were performed as described in the legend to Fig. 2, each employing a different competitor DNA ranging in concentration from 0.05- to 100-fold molar excess. The EFI protein-DNA complexes were excised from the dried gel and radioactivity in the complexes was quantitated by liquid scintillation counting. The ratio of DNA bound in the presence versus absence of competitor DNA for each concentration of competitor DNA added to the binding reaction was calculated and plotted as shown. The values for each point shown are the average of duplicate determinations. ⋄, EFI; ■, competitor DNA as indicated in the upper right corner of each graph. Nucleotide sequences of the EFI and competitor oligonucleotides used are given in Table I. ●, EFI point mutant at -129 (see Fig. 2); △, EFI LS9 linker insertion mutant (see Fig. 1).

EFI would recognize with equivalent or higher affinity the binding sites for these two CCAAT factors, as the overall sequence homology between their respective binding sites and that of EFI outside the CCAAT box is only 33% for NF-Y and 48% for CBF. We note that although EFI was capable of recognizing both a CP1-specific (i.e. a CCAAT binding site not recognized by CP2) and a CP2-specific (i.e. a CCAAT binding site not recognized by CP1) binding site with reasonably high affinity as well, EFI cannot be identical to either the CP1 or CP2 CCAAT factors identified in HeLa cells, since neither CP1 nor CP2 have been reported to cross-bind to the CCAAT boxes in the α-globin gene and H-2Kb gene at position −72 (Chodosh et al., 1988). The c/EBP and CTF/NF-I binding sites exhibited the weakest affinity for EFI of those CCAAT box-containing sites which we tested, although a 10-fold decrease in affinity might still be considered reasonable binding. Nonetheless, because EFI is not heat-stable, we do not think it is related to the CCAAT/GCAAT factor c/EBP (Graves et al., 1986; Johnson et al., 1987). Also, as shown below, 32P-CTF/NF-I gives a different mobility shift pattern than 32P-EFI DNA. Combined with additional observations described below, we argue that EFI is a CCAAT binding factor distinct from CTF/NF-I.

To further explore the relationship between EFI, CBF, and NF-Y, we radiolabeled the synthetic oligonucleotides representing each of these NF-Y factor binding sites as well as those for CP1, CP2, and CTF/NF-I for use in direct binding experiments with the chick embryo nuclear extract. As the autoradiograph in Fig. 5 (lanes 1–6) shows, a retarded band of identical mobility is formed with 32P-EFI, 32P-NF-Y, 32P-CBF, and 32P-CP2 DNAs. In contrast, the 32P-CTF/NF-I DNA complexes in nuclear extracts from different cell types. 1 ng of 26–30-bp 32P-labeled oligonucleotide representing the binding sites for EFI, CBF, NF-Y, CP2, CP1, or CTF/NF-I (sequences of the oligonucleotides are given in Table I) was mixed in the presence of poly(dI)-poly(dC) with nuclear extract prepared as described under "Materials and Methods" from chick embryos (lanes 1–6), the rat hepatocyte H4 cell line (lanes 7–12) or the P388 mouse macrophage cell line (lanes 13–18). In lanes 19–24, the radiolabeled DNAs were mixed with a whole cell transcription extract prepared from HeLa cells as described by Shapiro et al. (1988). For a given extract, the same quantity of protein was incubated with each of the DNAs; however, the amount of protein in the binding reactions varied among the different extracts, ranging from 2 to 5 μg. The samples were analyzed by electrophoresis and autoradiography of the dried gel.

FIG. 5. Electrophoretic analysis of CCAAT factor protein-DNA complexes in nuclear extracts from different cell types.
DNA gives a smeared retarded band pattern that is clearly distinct from the discrete EFI band. Incubation of $^{32}$P-labeled CP1 DNA with the chick embryo nuclear extract resulted in a hybrid pattern; both the smear typical of CTF/NF-I and the discrete band characteristic of EFI/NF-Y/CP2/CFB were observed. This is consistent with the weaker affinity of the CP1 DNA for EFI in competition experiments, as well as the fact that the $\alpha$-globin CCAAT box is also a binding site for EFI. A second poly(dI)-poly(dC)-resistant complex is formed with the $^{32}$P-labeled CBF DNA. This complex is also observed with the $^{32}$P-EFI DNA upon longer exposure. Because of its faster mobility this complex may consist of a proteolytic product of EFI present in chick embryo nuclear extracts which still retains the ability to specifically bind to the highest affinity EFI recognition sequences.

We also examined the mobility shift patterns of radiolabeled EFI, IF, NF-Y, CP2, CBF, and CFT/NF-I DNAs in extracts prepared from different species and cell types previously employed in the characterization of these different CCAAT-binding proteins. In particular, we prepared extracts from rat hepatocytes, where CBF has been identified (Hata-mochi et al., 1988), and from the P322 mouse macrophage cell line known to contain NF-Y (Hooft van Huijsduijnen et al., 1987). We also prepared a transcription extract from HeLa cells where CP1, CP2, and CFT/NF-I (Chodosh et al., 1988) have been characterized. As shown in Fig. 5 (lanes 7–24), incubation of each of these nuclear or transcription extracts with the radiolabeled EFI, NF-Y, CBF, and CP2 oligonucleotides again resulted in the formation of the characteristic discrete retarded EFI protein-DNA complex. The radiolabeled CTF/NF-I oligonucleotide exhibited a smeared pattern in each of the extracts, and both the discrete band and the smeared pattern formed with the radiolabeled CP1 oligonucleotide to varying extents depending on the extract. The HeLa transcription extract contains a second, more rapidly moving band which formed with all the CCAAT box oligonucleotides. This band is generated by a nonspecific protein-DNA interaction which is not blocked by the levels of poly(dI)-poly(dC) used here, but which can be eliminated using higher levels of nonspecific DNA.

In their analysis of the CCAAT-binding proteins present in HeLa cells, Chodosh et al. (1988) did not report on a binding activity which could recognize both the CCAAT box in the $\alpha$-globin gene (CP1 DNA) and the CCAAT box in the H-2K$^b$ gene at position –72 (CP2 DNA). However, as shown in Fig. 6, the protein in the HeLa transcription extract which binds to $^{32}$P-EFI DNA is competed by both CP1 and CP2 DNAs. We thus conclude that a fourth CCAAT-binding protein exists in HeLa cells, which we will refer to as EFI. Quantitation of the competition experiments presented in Fig. 6 indicates that HeLa EFI exhibits an approximately 2-fold lower affinity for the CP2 DNA and a 10-fold lower affinity for the CP1 DNA, relative to the high affinity EFI binding site in the RSV LTR.

These values are similar to those observed for EFI in the chick embryo nuclear extracts, and appear to distinguish the EFI CCAAT factor from either the CP1 or CP2 CCAAT proteins characterized by Chodosh et al. (1988). Although we refer to this fourth CCAAT-binding protein in HeLa cells as EFI, the observation that EFI in chick embryo extracts (and also in HeLa, rat hepatocyte, and mouse macrophage extracts (data not shown)) binds with equivalent or greater affinity to the NF-Y and CBF binding sites and the observation that identical mobility shift patterns are observed with radiolabeled EFI, NF-Y, and CBF DNAs, not only in the chick embryo extract, but also in rat hepatocyte, mouse macrophage, and HeLa extracts, suggest that EFI, NF-Y, and CBF may in fact be the same or a highly related ubiquitous CCAAT DNA binding protein. Of course, it is possible that different proteins could give rise to the same retarded band in mobility shift assays, and thus we have pursued the purification of EFI, to compare its biochemical properties with those reported for CBF and NF-Y.

Ion Exchange Chromatography of EFI: Comparison with CBF and NF-Y—The biochemical fractionation of CBF and a factor which binds to the Y box of the major histocompatibility class II gene I-Aβ (NF-Y?) have been previously reported (Hatamochi et al., 1988; Celada and Maki, 1989). Studies of this nature have demonstrated that these CCAAT factors are actually composed of two different components, both of which are required for DNA binding. In particular, ion exchange chromatography (MonoQ) of CBF and the factor which binds to the Y box of the major histocompatibility class II gene I-Aβ (NF-Y?) required the addition of bound material (factor A) to the column flow-through (factor B) to recover substantial DNA binding activity. To compare the behavior of EFI to CBF (and possibly NF-Y?), an EFI-containing chick embryo nuclear extract in 5% glycerol was prepared and fractionated on the anion exchanger Q-Sepharose. After collection of the flow-through and bound-column material, individual or mixed fractions were assayed for DNA binding activity either in 5 or 25% glycerol (Fig. 7). Whereas the unfractionated nuclear extract contained competent EFI, very little EFI binding activity was recovered in either the column flow-through or the bound material. Upon mixing the flow-through and 1 M NaCl eluate, we observed a great increase in the EFI DNA binding activity corresponding to 20% of the original, however, only when the binding assay was performed in 25% glycerol (Fig. 7, compare lanes 4 and 8). When only 5% glycerol was present, no significant increase was observed. After the salt was decreased in the eluate by dialysis, an even greater portion of the applied EFI DNA binding activity was recovered upon mixing, 40–50% (data not shown). These results are essentially the same as those observed by Hata-mochi et al. (1988). Although perhaps not intentionally, they also assayed for CCAAT box-binding activity under high glycerol conditions by adding a loading buffer containing 30%...
Additional complexes present are due to nonspecific protein-DNA resistance. The position of the EFI protein-DNA complex is indicated.

were immediately assayed either individually (lanes 2, 4, 6, and 7) or mixed (lanes 4 and 8) for EFI DNA binding activity with the radiolabeled [3H] and was not adjusted prior to native acrylamide gel electrophoresis. This reaction was collected as described under "Materials and Methods." The mobility samples was either 5% (lanes 1-4) or 25% (lanes 5-8) for EFI DNA binding activity with the radiolabeled 41-bp EFI oligonucleotide. The final glycerol concentration in the gel mobility samples was either 5% (lanes 1-4) or 25% (lanes 5-8) and was not adjusted prior to native acrylamide gel electrophoresis. The position of the EFI protein-DNA complex is indicated. Additional complexes present are due to nonspecific protein-DNA interactions as no poly(dI)-poly(dC) was included in these binding reactions.

glycerol to their samples before gel electrophoresis. This interpretation assumes that the effect of glycerol is essentially instantaneous. We have tested this and found it to be so (data not shown). In both cases the high glycerol concentration did not increase the EFI or CBF activity in the two fractions unless they were mixed. The effect of glycerol on EFI and CBF activity will be discussed in more detail elsewhere (Ozer, 1990). Binding of the Q-Sepharose fractions to the radiolabeled CCAAT oligonucleotides representing binding sites for CBF, CP2, CP1, CTF/NF-I, and NF-Y was also examined (Fig. 8). The column flow-through (lanes 1-6) is nearly void of EFI, NF-Y, CP2, and CBF binding activity and only a small amount of DNA binding activity specific for these oligonucleotides was detected in the elution (lanes 7-12). When the column flow-through and elution were mixed there was a 12-18-fold increase in binding to the EFI, NF-Y, CP2, and CBF oligonucleotides (lanes 13-18). In contrast, greater than 90% of the detectable CTF/NF-I and CP1 activity was present in the flow-through alone. Mixing led to no or only a modest increase (3-fold) in these activities, respectively.

The stability of the two Q-Sepharose fractions to heating was tested by incubating either the dialyzed flow-through or column elution for 5 min at 60°C. The heat-treated component was then tested for DNA binding activity either alone or mixed with the appropriate untreated complementing factor. Hatamochi et al. (1988), found that the component present in the column flow-through (factor B) was stable, whereas the factor which bound the Q column (factor A) was inactivated by the 5-min incubation. Our results were identical (data not shown).

Molecular Weight Analysis of EFI and CBF—The molecular weights of the two components which bind to the quaternary amine resin (factor A) or flow-through (factor B) have been estimated for CBF and the NF-Y-like factor by renaturation of DNA binding activity after SDS-gel electrophoretic analysis. Factor A was estimated to have an approximate molecular weight of 34,000 (Celada and Maki, 1989) or 39,000 (Hatamochi et al., 1988), whereas factor B was estimated to have a molecular weight of 41,000 (Hatamochi et al., 1988) or 42,000-46,000 (Celada and Maki, 1989).

We performed a similar molecular weight analysis with EFI present in the chick embryo nuclear extract. A sample of extract was analyzed by SDS-polyacrylamide gel electrophoresis, transferred to Immobilon paper, and the paper was cut into slices. Proteins were eluted from each slice, renatured, and tested for EFI DNA binding activity in the absence or presence of factor A (Q-Sepharose bound) or factor B (heat-treated chick embryo extract). In the absence of exogenous factor A or factor B, a small amount of EFI DNA binding activity was recovered in fraction 6 (presumably due to the overlap of factors A and B in this fraction). However, a substantial increase in EFI DNA binding activity was observed when the SDS-gel fractions were complemented with either factor A or factor B. Upon addition of factor B (Fig. 9A), the peak of EFI DNA binding activity was observed in fractions 5 and 6, suggesting that factor A migrated in the SDS-gel with a molecular weight of approximately 43,000–60,000. Upon complementation with factor A (Fig. 9B), the peak of EFI DNA binding activity was observed in fraction 6, corresponding to a molecular weight of 41,000-48,000 for factor B.
EFI Is Highly Degenerate in Sequence Recognition Outside the CCAAT box. Clearly, sequences outside the inverted CCAAT box determine the binding affinity for EFI, although these additional determinants of EFI binding appear to be highly degenerate. In addition to the 8902 and wild type/pol mutants, we have identified five other sequences to which EFI binds with high affinity in our studies presented here. These are the CBF, NF-Y, CP2, and CP1 binding sites, as well as the second EFI binding site in U3 at -82 to -57. Only one of these sites exhibits greater than 50% homology with the high affinity EFI binding site at -121 to -146 in the RSV LTR (the CP1 site has 15 out of 26 matches). Furthermore, no single nucleotide in the sequences outside the inverted CCAAT box is conserved among all of these binding sites. A consensus EFI recognition sequence compiled from these six high affinity binding sites for EFI is as follows.

AAAG CCAG CGAA A
GCCCGAGGAAGAGGAGG
CGTCGTGTTCC

Of the 21 nucleotides outside of the inverted CCAAT box, all four nucleotides (X) are allowed at seven positions, three of four nucleotides at 12 positions, in only two positions are just two nucleotides found, and in only one position is a restriction to pyrimidines (Y) noted. It appears that although additional protein-DNA contacts outside of the inverted CCAAT box are required for high affinity binding, no specific protein-DNA contact is obligatory.

Extreme degeneracy in DNA sequence recognition by eukaryotic trans-acting factors has been previously noted among several other transcriptional activator proteins. One of the most well characterized examples is the yeast HAP1 activator. HAP1 binds to both the upstream activator sequence US1 of the CYC1 gene and the UAS of the CYC7 gene, even though the UAS sequences of these respective genes share only seven matches in a 23-bp region (Pfeifer et al., 1987). Baumraker et al. (1988) have presented an extensive characterization of the remarkably flexible DNA sequence recognition by the ubiquitous 100,000-dalton HeLa cell octomer binding protein OBPI0 (also known as NF-A1, NFIII). Although this protein normally binds to the highly conserved octomer motif ATGCNAAAT present in the immunoglobulin promoter and enhancer regions as well as in various other gene promoters, the ubiquitous octomer factor can also recognize degenerate octomer motifs, given appropriate compensatory sequences in the flanking DNA. We have not investigated whether EFI may also tolerate variations in the CCAAT motif, given appropriate flanking sequences to provide compensatory protein-DNA interactions.

EFI = CBF = NF-Y?—Since different combinations of protein-DNA contacts can apparently lead to high affinity EFI binding, identifying these possible contacts through site-directed mutagenesis will probably be very difficult, due to the potential number of different sequence variations one must test. In the absence of a consensus recognition sequence (other than CCAAT) for EFI binding, we resorted to directly assaying whether EFI would bind to the recognition sites for other previously described CCAAT transcription factors. Although the overall homologies between the high affinity EFI binding site in the RSV LTR and the other CCAAT factor binding sites which we tested were (except for CBF) less than 50%, and less than 33% outside of the CCAAT box, we nonetheless found that EFI would recognize all of the CCAAT factor binding sites we tested with reasonable affinity. EFI in our avian nuclear extracts showed a 10-fold lower affinity for the c/EBP and CTF/NF-I binding sites relative to the EFI (-129) binding site in competition studies. Because EFI is not heat-stable, it is clearly distinct from c/EBP (Johnson et al., 1987). Moreover, we have shown that EFI is a two-
component CCAAT factor, whereas o/EBP and CTF/NF-I binding activities reside in single polypeptides (Landschulz et al., 1988a; Santoro et al., 1988). This fact, coupled with the different mobility shift patterns of 32P-PEFI and 32P-CTF/NF-I DNAs in extracts from chick embryo, rat hepatocyte, mouse macrophage, and HeLa cells, establish that EFI is not the CCAAT-binding protein CTF/NF-I.

A number of two-component CCAAT binding proteins have been described. Hatamochi et al. (1988), identified a two-component CCAAT specific factor, CBF, in extracts of rat liver or NIH 3T3 cells which binds to the promoter of the α2(I)-collagen gene. Celada and Maki (1989), have characterized a two-component CCAAT factor which binds to the Y box of the major histocompatibility class II gene I-Aβ. In HeLa cells, Chodosh et al. (1988) reported on the existence of three distinct CCAAT-binding proteins, CP1, CP2, and NF-I, each composed of two different types of subunits. These three CCAAT binding proteins were recognized as distinct because, in addition to different methylation interference patterns and chromatographic behavior, at least some high affinity recognition sites for each of these proteins were not recognized by the other two proteins. We have found that the EFI DNA binding activity in chick embryo extracts and in a HeLa whole cell transcription extract can recognize with high affinity the CP2-specific site at −72 in the H-2Kk major histocompatibility gene promoter and with only 6- to 10-fold lower affinity, the CP1-specific site in the α-globin gene promoter. This EFI (and CBF, NF-Y) DNA binding activity therefore represents a four-component CCAAT binding protein present in HeLa cells. We note that Chodosh et al. (1988), used partially purified chromographic fractions from HeLa whole cell transcription extracts that had been passed over several different columns before recombining. Our experiments were performed with unfractionated HeLa whole cell extract. Perhaps in the chromatographic fractionation, the heterologous subunits of EFI were separated into different fractions than those recombined by Chodosh et al. (1988) in their studies.

Although EFI appears to be distinct from CP2 and CP1, we wish to point out a number of similarities between EFI and CP1, which suggest that EFI is more highly related to CP1 than to CP2. Of the six high affinity binding sites for CP1 reported by Chodosh et al. (1988), we have identified four sites (human α-globin; H-2Kk (−51), a Y box; adenovirus major late promoter; UAS9 [1P1-CYC1] (Fig. 4 and Greuel et al., 1990)) and found all to be high affinity binding sites for EFI as well. Both CCAAT binding proteins CP1 and EFI recognize the CCAAT box in the herpes simplex virus thymidine kinase promoter with lower affinity (Fig. 4 and Chodosh et al., 1988). In contrast, CP2 recognizes with high affinity a CCACT motif present in the γ-fibrinogen gene promoter (Chodosh et al., 1988). We have shown that this A to C mutation in the CCAAT motif of the high affinity EFI binding site at −132 in the RSV LTR reduces EFI binding by 60-fold (Fig. 2), and thus, EFI would be unlikely to bind to the γ-fibrinogen CP2 binding site. The methylation interference patterns of CP2, CP1, and EFI on the γ-fibrinogen, adenovirus major late promoter, and high affinity RSV LTR binding sites, respectively, have been reported (Chodosh et al., 1988; Greuel et al., 1990). The CP2 and EFI methylation interference patterns are very different, whereas the CP1 and EFI patterns are very similar as shown in Table II. In fact, the only indication that CP1 and EFI are different CCAAT binding proteins is the observation that CP1 does not bind to the CCAAT box at −72 in the H-2Kk gene promoter (affinity reduced three orders of magnitude, Chodosh et al., 1988) whereas EFI binds to this CP2-specific site with an affinity approximately 3-5-fold higher than a CP1-specific site, the CCAAT box in the human α-globin gene promoter, when either a chick embryo fibroblast extract or HeLa whole cell transcription extract are tested.

We find no major differences between the two-component CCAAT binding proteins, CBF and EFI. Indeed, Hatamochi et al. (1988), have reported that CBF binds to the RSV LTR at both the −129 to −133 inverted CCAAT box and the −65 to −69 inverted CCAAT box recognized by EFI. We show here that EFI fractionates on the anion exchanger Q-Sepharose in the same manner as CBF upon MonoQ chromatography. A heat-sensitive bound fraction (EFla, CBFa) and a heat-stable flow-through fraction (EFIB, CBFb) must be recombined to recover substantial DNA binding activity. Interestingly, we found that mixing was effective only when the glycerol concentration of the samples was elevated (25%). We have found that EFI DNA binding activity in chick embryo extracts is very unstable upon purification by oligonucleotide affinity chromatography, and we have been unable to purify this multisubunit protein in homogeneity. We have, however, estimated the apparent molecular weight of EFIa and EFIb by SDS-polyacrylamide gel electrophoresis of a crude nuclear extract followed by elution of proteins from individual gel slices. When complemented with EFIa and EFIb, respectively, we estimated EFIa to have a M, of 40,000-60,000, and EFIb, a M, of 41,000-46,000. These estimates for EFIb and EFIa are in good agreement with that reported by Hatamochi et al. (1988) for CBFb (41,000). Our molecular weight estimate for EFIa is larger than that reported for CBFa, 39,000, Hatamochi et al., 1988). Our analysis was performed with unfractionated nuclear extract, whereas Hatamochi et al. (1988) utilized material from an oligonucleotide affinity column. We have found, in our efforts to purify EFI, that EFI is particularly unstable, presumably due to proteolysis. Thus, the variance in molecular weight estimates for factor A may reflect differences in degree of proteolysis.

No comparable fraction of NF-Y by ion exchange chromatography has been reported to allow us to compare the behavior of EFI with this Y box-binding CCAAT factor. However, a CCAAT-binding factor which recognizes the Y box of the MHC class II gene I-Aβ was partially purified from a mouse B-cell lymphoma line and shown to fractionate in an identical manner to CBF and EFI on a MonoQ resin (Celada and Maki, 1989). The column bound (factor A) and column flow-through (factor B) components required for specific DNA binding activity were estimated to have molecular weights of 34,000 and 42,000-46,000, respectively, upon elution from an SDS-polyacrylamide gel (Celada and Maki, 1989). These values are similar to those reported for CBF, and as noted above, EFI. Methyltransferase interference experiments have been performed for NF-Y binding to the inverted CCAAT box (Y box) of the murine MHC class II E gene (Dorn et al., 1987). The methylation interference pattern for NF-Y, shown in comparison with the patterns reported for CP1 and EFI in Table II, are again very similar for the three CCAAT-binding factors. Cross-competition experiments with radiolabeled E, Y box DNA and α-globin or thymidine kinase CCAAT box competitor DNAs performed by Dorn et al. (1987), have shown that the relative affinities of NF-Y for these DNAs follow the order NF-Y > α-globin > thymidine kinase with the thymidine kinase DNA exhibiting an approximately 10-fold lower affinity. This is in very good agreement with our results examining EFI binding to these DNAs in chick embryo extracts (Fig. 4). Finally, Didier et al. (1988), have reported on the isolation of a human cDNA for the Y box factor, YB-1, which binds to
the human HLA-DR α chain gene. As will be reported elsewhere (Ozer et al., 1990), we have recently isolated a cDNA for EFIα from a rat liver expression library. The amino acid sequence of the proteins encoded by these two cDNAs are >95% identical. It may seem paradoxical that both we and Didier et al. (1988) obtained these respective cDNA clones by screening an expression library; such an approach would not, a priori, be expected to be successful if two different polypeptide components are required for specific DNA binding. However, during the course of our biochemical fractionation studies with EFI, we made some observations concerning the effect of glycerol on EFI binding activity which suggests that two forms of the protein may exist. One is the heterologous multisubunit factor composed of EFI1 and EFI2 polypeptides and the second, a homologous, perhaps multisubunit, factor composed of only EFI1 polypeptides which also appears to be capable of specific DNA recognition. Further characterization of this second form of EFI will be presented elsewhere (Ozer et al., 1990); however, since specific DNA binding is dependent on a single polypeptide, a cDNA for such a CCAAT binding factor could be identified by expression library screening.

Functional significance of EFI/CPF/NF-Y—Although in the work presented here we have not considered the functional significance of EFI for LTR enhancer/promoter function, we and others (Dutta et al., 1990; Greuel et al., 1990) have previously reported that disruption of either the CCAAT box at position −129 or that at position −65 has deleterious effects on the transcriptional activity of the RSV LTR in vivo. In fact, linker substitutions and point mutations within the high affinity EFI binding site at −121 to −146 which have the most drastic effects on EFI binding in vitro, also exhibit the most severe reduction in transcriptional activity from the RSV LTR in vivo (Greuel et al., 1990), suggesting that EFI plays an important role in LTR enhancer/promoter function. It has recently been reported (Dutta et al., 1990) that the DNA binding activity of EFI is induced by serum and the oncogene v-src. It has been known for some time that cells transformed by Rous sarcoma virus exhibit changes in gene expression (Shui et al., 1977; Adams et al., 1979; Fagan et al., 1981; Hendriks and Weintraub, 1984; Groudine and Weintraub, 1988; Matrissian et al., 1988; Gal and Gottesman, 1986; Sugano et al., 1987; Jones and Luk, 1988; Bell et al., 1990). The finding that EFI DNA binding activity is elevated by v-src suggests that transformation by RSV results in a major reduction in the level of RNA encoding a2(1)-collagen in chick embryo fibroblasts (Adams et al., 1979). Since CBF, the CCAAT-binding protein which binds to the promoter of this gene, and EFI appear to be the same or a highly related transcription factor, it is possible that v-src induction of EFI/CFB binding activity could have negative, as well as positive, influences on gene transcription. Given that EFI and NF-Y also appear to be the same or a highly related transcription factor, it would be interesting to determine if v-src, acting through EFI/NF-Y, were to also influence MHC class II gene expression. The availability of cDNA clones for this highly related or identical group of CCAAT-binding proteins, EFI/NF-Y(YB-1)/CFB, should facilitate further investigation into the role of this ubiquitous factor in positively (or negatively?) influencing transcription initiation, both constitutively, and in response to serum and oncogenes.

Acknowledgments—We thank our colleagues in the Sealy and Chalkley laboratories for their support, gifts of extracts and oligonucleotides, and for critical reading of the manuscript. We are indebted to Ginger Griffis and Kathy Mattix for providing expert technical assistance. We also acknowledge the Diabetes Research and Training Center Molecular Biology Core at Vanderbilt University for oligonucleotide synthesis.

REFERENCES

Jones, S. W., and Luk, K.-C. (1988) J. Biol. Chem. 263, 9607-9611