Enhancer factor I (EFI) is a trans-acting factor which binds to the Rous sarcoma virus long terminal repeat enhancer and promoter at two inverted CCAAT-box motifs. We demonstrate that two forms of EFI DNA binding activity exist in nuclear extracts of avian cells. One form requires two heterologous components (EFI/A/EFI/B) for high affinity, specific DNA binding activity, whereas a second form is not dependent on EFI, for binding and may be composed solely of EFI, perhaps as a multimer. Both forms give rise to the same mobility shift in gel retardation assays, but the two forms can be separated chromatographically under buffer conditions which stabilize the two DNA binding activities. A cDNA for EFI, has been isolated from a rat liver cDNA expression library. The 1489-base pair EFI cDNA encodes a 322-amino acid protein which is nearly identical to two previously described human DNA binding proteins. These are dbpB, a DNA binding protein of unknown specificity which binds to the epidermal growth factor receptor enhancer and c-erbB-2 gene promoter (Sakura, H., Maekawa, T., Imamoto, F., Yasuda, K., and Ishii, S. (1988) Gene (Amst.) 73, 499-507), and YB-1, a protein which recognizes the Y-box (inverted CCAAT motif) of the HLA-DR (YB-1) gene (Didier, D. K., Schiffenbauer, J., Woulfe, S. L., Zacheis, M., and Schwartz, B. D. (1988) Proc. Natl. Acad. Sci. U. S. A., in press). This highly conserved domain among all six proteins is presumed to the GenBank/EMBL Data Bank with accession number(s) J05704.

Sequence-specific DNA binding proteins comprise a large class of regulatory proteins which play pivotal roles in modulating gene expression in eukaryotic cells. The cis-elements with which they interact in regulating transcription have been broadly classified into two groups, promoter proximal elements and enhancers (Dyan, 1989; Dynan and Tjian, 1985; Mitchell and Tjian, 1989). Promoter proximal elements are located near to the start site of transcription, whereas enhancers function near independently of position, orientation, and distance with respect to the site of transcription initiation (Dynan, 1989; Dynan and Tjian, 1985; Ptashne, 1986; Serfling et al., 1985). The sequence, 5'-CCAAT-3', is a frequent motif in the promoter proximal regions of a number of promoters. Deletions or point mutations in this motif have been shown to detrimentally affect levels of transcription from a variety of promoters including, for example, the human a-globin (Mellon et al., 1981), mouse b-globin (Myers et al., 1986), a(2)I collagen (Karsenty et al., 1988), RSV LTR (Dutta et al., 1990; Greuel et al., 1990), human hsp70 (Morgan et al., 1987), and herpes simplex virus tk promoters (Graves et al., 1988). CCAAT-box binding proteins have been described which bind to these and a variety of other promoters in nuclear extracts from different cell types (Barberis et al., 1987; Celada and Maki, 1989; Chodosh et al., 1988; Cohen et al., 1986; Dorn et al., 1987; Gallinari et al., 1989; Goding et al., 1987; Graves et al., 1987; Hatamochi et al., 1988; Jones et al., 1985, 1987; Knight et al., 1987; Morgan et al., 1987; Olesen et al., 1987), cDNA clones have been isolated and characterized for three different CCAAT binding proteins, CTF/NFI from human HeLa cells (Mermod et al., 1989; Santoro et al., 1988), C/EBP from rat liver (Landschulz et al., 1988a, 1988b), and YB-1, a protein which binds to the Y-box (inverted CCAAT box) in the promoter of the human HLA-DR 1-a chain gene (Didier et al., 1988). In contrast, a subset of the CCAAT binding factors have been described which appear to be composed of two heterologous components, both of which are required for specific DNA binding activity. These include the yeast factors HAP2/HAP3 (Olesen et al., 1987), as well as CBF (Hatamochi et al., 1988), which binds to the promoter of the a(2)I collagen gene, a protein which binds to the Y-box of the MHC class II I-A(2) gene promoter (Celada and Maki, 1989), and the CCAAT factors CP1, CP2, and NFI identified by Chodosh et al. (1988) in HeLa cells. Although the yeast genes encoding HAP2 and HAP3 have been well characterized (Hahn et al., 1988; Pinkhan and Guarente, 1985), the molecular structure of the two component CCAAT factors from...
higher eukaryotes have yet to be determined.

We have focused our attention on a factor first identified in avian nuclear extracts which binds to two inverted CCAAT motifs in the RSV LTR enhancer and promoter (Sealy and Chalkley, 1987). Deletions of either CCAAT motif reduce the transcriptional activating ability of the RSV LTR enhancer and promoter to less than 5% of wild type levels (Datta et al., 1990; Greuel et al., 1990). The factor which binds to these inverted CCAAT motifs in vitro, enhancer factor I (EFI), appears to belong to the latter class of CCAAT factors described above, requiring two heterologous components for specific DNA binding activity. EFI DNA binding activity has been identified in a variety of different cell types. In fact, we have argued based on the high affinity of EFI for the binding sites of CBF and NF-Y, a protein which binds to the Y-box in the promoter of the mouse MHC class II Eo gene, as well as biochemical fractionation, methylation interference patterns, and molecular weight analysis, that EFI is identical or highly related to CBF and the group of Y-box binding proteins (Faber and Sealy, 1990).

We report here on the isolation of a cDNA clone, from a rat liver cDNA expression library, encoding one of the two heterologous components of EFI, EFI*. This clone was obtained by screening an expression library with a radiolabeled high affinity EFI recognition sequence. This may seem paradoxical, as such an approach would not a priori be expected to be successful if two different peptide components are required for specific DNA binding. However, we have found that under the appropriate buffer conditions, EFI, can bind to DNA specifically in the absence of EFL in vitro. The amino acid sequence of EFI, is 95% homologous to the protein encoded by the human cDNA for YB-1 and 97% homologous to a protein with unknown DNA binding specificity encoded by the human cDNA, dbpB. Analysis of the molecular structure of EFI, reveals a novel arrangement of alternating clusters of positive and negative charges not yet reported for any anti-acting factor. As discussed, the functional significance of this novel structural motif is supported by the finding that the alternating clusters of charged residues are maintained in the cDNAs encoding other CCAAT binding proteins. Possible functions for this novel structural motif will be discussed.

MATERIALS AND METHODS

Extract Preparation

Nuclear extracts from 12-14-day-old chick embryos were prepared as previously described (Boulden and Sealy, 1990) except that 10 mM benzamidine was included in all buffers. A 0.1 M NaCl (Buffer B) extract of chick embryo nuclei was used for all experiments presented here. Extracts from bacterial cells were prepared by the method described in Samбрenk et al. (1986) adapted from Marston (1987), except that 3 mM MgCl2 was added to the lysate for DNase I treatment. EFI, DNA binding activity was found in the soluble fraction. For some experiments, bacterial extracts were prepared using polyethylene glycol (PEG) precipitation as described by Gross et al. (1976) with the following modifications. Solution A contained 10 mM Hepes, pH 8, 25% glycerol, and 100 mM NaCl. Solution C contained 25% glycerol in addition to 0.02 M EDTA and 0.08% (w/v) deoxycholate, while 1 M NaCl was omitted. Solution E contained 5% PEG (w/v), 2 M NaCl, 10 mM Hepes, pH 8, and 25% glycerol. Dithiothreitol was not added to buffers but the following protease inhibitors were added just before use: 1 mM phenylmethylsulfonyl fluoride, 5 mg/liter leupeptin, 0.1 mM pepstatin, and 10 mM benzamidine. Extraction of proteins from the PEG precipitate was performed in twice the recommended volume and the resulting supernatant was dialyzed to 10 mM Hepes, pH 8, 1 mM EDTA, 50 mM NaCl, 5% glycerol, 10 mM benzamidine, and 0.1 mM phenylmethylsulfonyl fluoride before ion exchange chromatography or to the same buffer except containing 50% glycerol for other applications.

Chick Embryo Nuclear Extract—An EFI-containing chick embryo nuclear extract was prepared and dialyzed to buffer Q (10 mM Hepes, pH 8, 0.1 mM EDTA, 50 mM NaCl, 7 mM β-mercaptoethanol, 10 mM benzamidine, and 0.1 mM phenylmethylsulfonyl fluoride) containing 50% glycerol (Faber and Sealy, 1990). Extract from 0.65 embryo was subjected to Q-Sepharose fractionation (0.2 ml of packed resin) essentially as described previously (Faber and Sealy, 1990), except that all buffers contained 50% instead of 5% glycerol. Only the 1 M NaCl fraction was dialyzed to buffer Q containing 50% glycerol before analysis by the gel electrophoretic mobility shift assay. Q-Sepharose fractionation of the EFI-containing chick embryo nuclear extract in 5% glycerol was performed as described previously (Faber and Sealy, 1990).

Bacterial EFI—Bacterial extracts were prepared by PEG precipitation as described above from BL21(DE3)pLysS cells transformed with either pET3b or pET3b+EFI, harvested 17 h after IPTG induction. Extracts from 15-ml cultures in 0.75-ml final volume were incubated for 6 h at 4°C with 0.5 ml (packed resin) of Q-Sepharose equilibrated in buffer Q with 5% glycerol. The unbound material was collected, the resins were washed with 0.75 ml of buffer Q containing 5% glycerol and the bound material eluted for 45 min in 0.75 ml of buffer Q containing 1.5 M NaCl and 5% glycerol. All samples were dialyzed to buffer Q containing 50% glycerol before analysis by the gel shift assay.

Gel Retardation Assay

Gel electrophoretic mobility shift assays were performed as described previously (Sealy and Chalkley, 1987; Faber and Sealy, 1990). Radiolabeled and competitor oligonucleotide DNAs were prepared as described previously (Boulden and Sealy, 1990; Faber and Sealy, 1990). In some instances the proportion of DNA bound was determined by excluding the protein-DNA complex from the dried gel and quantitating the radioactivity present by liquid scintillation counting.

Molecular Weight Determination of Bacterial EFI*

Bacterial extracts were prepared by the PEG precipitation method described above from cells transformed with pET3b or pET3b+EFI, harvested 17 h after IPTG induction. The extracts were subjected to 12% SDS-polyacrylamide gel electrophoresis (without heating), and transferred to Immobilon membrane (Immobion-P; Millipore). The membrane was cut into slices and the protein eluted and renatured as described previously (Faber and Sealy, 1990). The fractions were analyzed for EFI, by the gel electrophoretic mobility shift assay in the presence of poly(dI-dC) or poly(dI-dC). The molecular weight of proteins present in each Immobilon slice was determined by comparison with the mobility of prestained molecular weight markers (Bethesda Research Laboratories).

Expression Library Screening

A rat liver λgt11 expression library (Mueckler and Pitot, 1985) containing 1–3 x 106 inserts with a size range of 0.5–5.0 kilobases was screened employing the filter hybridization protocol essentially as described by Vinson et al. (1988). The radiolabeled DNA probe was generated by treating 800 ng of 41-bp duplex synthetic oligonucleotide, purified, and annealed as described previously (Boulden and Sealy, 1990), with T4 polynucleotide kinase (15 units, New England Biolabs) and [γ-32P]ATP (6000 Ci/mmol, Du Pont) in a final volume of 40 μl containing 0.2 M Tris, pH 8, 10 mM MgCl2, 15 mM dithiothreitol, and 5% glycerol. The 41-bp synthetic DNA was ligated by treatment with T4 DNA ligase (800 units, New England Biolabs) according to the manufacturer's specifications. The ligation mixture was incubated with bacterial λgt11 cosmid DNA to provide a source of DNA ligase (800 units, New England Biolabs) according to the manufacturer's specifications. The primary screen of the λgt11 library was plated at a density of 40,000 plaque-forming units/145 cm2. After 4 h at 42°C, plates were overlaid with 10 mM impropregnated IPTG (BRL) nitrocellulose filters (Millipore) for 6 h at 37°C to induce cDNA expression. Dried filters, with bound protein, were denatured in 6 M guanidine hydrochloride (Sigma) in binding buffer (25 mM NaCl, 5 mM MgCl2, 10 mM β-mercaptoethanol, and 25 mM Hepes, pH 7.9). Denaturation agent was diluted from filters with binding buffer in a stepwise fashion.
Filters were blocked with binding buffer containing 5% Carnation non-fat dry milk. Filters, transferred to a 0.25% non-fat dry milk binding buffer solution containing 1 μg/ml of single-stranded salmon sperm DNA, were probed with 10^5 to 10^6 cpm of ligated EFI 41-bp probe/ml binding buffer at room temperature. Subsequently, filters were quickly washed 3 times for 10 min in binding buffer containing 0.25% non-fat dry milk at room temperature before autoradiography.

Expression of EFIα in Bacteria

An EFIα expression construct was prepared by partial EcoRI digestion of the EFIα/Agt11 clone and preparative agarose gel electrophoresis to isolate the 1489-bp EFIα cDNA fragment. The vector pET3b (Studier et al., 1990) was digested with BamHI and the protruding ends of both the linearized vector and the 1489-bp EcoRI fragment were filled in with DNA polymerase I at 4 °C (Sealy and Chalkley, 1987) prior to ligation of the two DNAs. An appropriate aliquot of the ligation mix was used to transform Escherichia coli DH5α (Bethesda Research Laboratories). Clones containing the EFIα cDNA insert in the appropriate orientation were subjected to dideoxy sequencing to insure fidelity at the vector-insert junction and conservation of the appropriate EFIα reading frame. Such a clone (pET3b+EFIα) was then used to transform E. coli BL21(DE3)pLysS. This host strain transformed with either pET3b+EFI or the vector alone (pET3b) was grown as described (Studier et al., 1990). Upon reaching an OD_600 of 0.5, cultures were induced with 0.4 mM IPTG for 15–17 h.

DNA Sequencing

The EFIα cDNA clone insert, isolated by EcoRI digestion (Promega) from purified Agt11/EFIα phage DNA, was subcloned into the pGEM-7Zf(+) plasmid vector (Promega). Double-stranded dideoxy DNA sequencing was performed on the EFIα/pGEM-7Zf(+) construct using modified T7 DNA Polymerase (Sequenase II) and [α-32P]dATP (3000 Ci/mmol, Amersham Corp.). Sequencing products were electrophoresed on either 6 or 9% polyacrylamide and 7 M urea gels at a constant power of 75 watts.

RNA Isolation and Northern Blot Analysis

Total RNA was isolated from confluent rat liver H4 cells as previously described (Chomczynski and Sacchi, 1987). RNA was electrophoresed on a 6.5% deionized formaldehyde and 1% agarose gel. RNA was recovered on the ion exchange resin, Q-Sepharose (Bethesda Research Laboratories). Clones containing the EFIα cDNA insert in the appropriate orientation were subjected to dideoxy sequencing to insure fidelity at the vector-insert junction and conservation of the appropriate EFIα reading frame. Such a clone (pET3b+EFIα) was then used to transform E. coli BL21(DE3)pLysS. This host strain transformed with either pET3b+EFI or the vector alone (pET3b) was grown as described (Studier et al., 1990). Upon reaching an OD_600 of 0.5, cultures were induced with 0.4 mM IPTG for 15–17 h.

Protein Identification Resource, were used to search the SwissProt, Genbank, EMBL, and NBRF Sequence databases for sequence homology to EFIα.

RESULTS

Heterologous and Homologous Multisubunit Forms of EFI—
Several different CCAAT binding proteins have been shown to be composed of two heterologous components (A and B) both of which are required for specific DNA binding activity (Celada and Maki, 1989; Chodosh et al., 1988; Hatamochi et al., 1988; Olesen et al., 1987). We have recently reported that EFI is such a factor (Faber and Sealy, 1990). When fractionated on the ion exchange resin, Q-Sepharose, EFI separates into a bound component (EFI*) and a component in the column flow-through (EFIf). Mixing of both components is required to regain substantial DNA binding activity (see Fig. 1, lanes 5–8), resulting in a 15-fold stimulation of EFI DNA binding activity. The behavior of EFI is analogous to that reported for two other CCAAT binding proteins, CBF (Hatamochi et al., 1988) and a factor which binds to the Y-box of the MHC class II I-Aα gene promoter (Celada and Maki, 1989), when fractionated on a Mono Q resin. In each of these studies, the protein extract was applied to the resin in a buffer containing 5% glycerol. In our attempts to purify EFI to homogeneity we have found that the factor is very unstable. However, we noted that recovery of EFI DNA binding activity was improved when the factor was maintained in buffers containing 50% glycerol. When we attempted to fractionate EFI on a Q-Sepharose column in the presence of 50% glycerol, we were surprised to find that mixing of column-bound (1 M NaCl elution) and flow-through fractions was not required to recover substantial DNA binding activity (see Fig. 1, lanes 2 and 3). Instead, approximately 50% of the input EFI DNA binding activity applied to the column was recovered in the flow-through fraction (Fig. 1, lane 2) and another 25% of the input activity could be eluted from the column with 1 M NaCl (Fig. 1, lane 3). Mixing of the bound and flow-through fractions resulted in a simple additive effect (Fig. 1, lane 4). Furthermore, reapplication of the fractions individually to the Q-Sepharose resin in 50% glycerol did not alter their chromatographic behavior (data not shown). This indicates that the recovery of binding activity in both the column-bound and flow-through fractions is apparently a reflection of two distinct forms of EFI DNA binding activity in the nuclear extract which differ in their interaction with the Q-Sepharose resin, rather than a consequence of overloading the resin. (High levels of glycerol (50%) did not restore binding activity to the individual fractions generated by Q-Sepharose chromatography in 5% glycerol (data not shown). Mixing following

FIG. 1. Identification of two forms of EFI binding activity. Nuclear extract from 14-day chick embryos (CE ex) in buffer Q containing 50% glycerol (lanes 1–4) or in buffer B containing 5% glycerol (lanes 5–8) was mixed with the ion exchange resin, Q-Sepharose, equilibrated in buffer Q containing 50 or 5% glycerol, respectively. The column flow-through (f) and a high salt elution were collected as described under “Materials and Methods” employing buffers containing 50 or 5% glycerol, respectively. The nuclear extract (lanes 1 and 5) and the Q-Sepharose column fractions were assayed either individually (lanes 2, 3, 6, and 7) or mixed (lanes 4 and 8) for EFI DNA binding activity with the radiolabeled 41-bp EFI oligonucleotide. The final glycerol concentration in all gel mobility assay samples was 25%. 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.
dialysis to 50% glycerol was still required to recover substantial EFI DNA binding activity.)

To determine the composition of the EFI DNA binding activities present in the column-bound and flow-through fractions obtained after passage over the Q-Sepharose resin in 50% glycerol, both fractions were diluted to 5% glycerol and passed again over the ion exchange resin. Material from the 50% glycerol column flow through behaved identically to unfractionated EFI when applied to the Q-Sepharose resin in 5% glycerol. As shown in Fig. 2, little binding activity was recovered in either the flow-through (lane 2) or column-bound (lane 3) fractions. Mixing led to a marked stimulation of specific EFI DNA binding activity (lane 4). By analogy to the behavior of unfractionated EFI, the 50% glycerol column flow-through contains both EFIα and EFIβ, which can be separated upon passage over Q-Sepharose, but only when the glycerol concentration is low (5%).

In contrast, when the 50% glycerol column-bound material was diluted to 5% glycerol and reapplied to Q-Sepharose, the column-bound fraction contained a small amount of EFI binding activity (Fig. 2, lane 7), whereas no EFI binding activity was detected in the flow-through (Fig. 2, lane 6). Mixing did not increase specific DNA binding activity (Fig. 2, lane 8). However, addition of a Q-Sepharose flow-through fraction known to contain EFIα (obtained by passage of an unfractionated extract over the resin in 5% glycerol) resulted in a marked stimulation of EFI DNA binding activity (Fig. 2, lane 9). These results suggest that the material which binds to Q-Sepharose in 50% glycerol (Fig. 1, lane 3) contains EFIα, but not EFIβ.

A summary of the results of fractionating an EFI-containing nuclear extract over the Q-Sepharose resin in either 50 or 5% glycerol is presented in Fig. 3. In buffers containing high glycerol, 50%, the EFI DNA binding activity which is not composed of both EFIα and EFIβ is depicted as (EFIα)α. Although we have not experimentally demonstrated that this DNA binding activity consists of multimers of the EFIα factor, the gel shift pattern produced (which is identical to that produced by the heterologous complex EFIα,EFIβ) indicates it consists of more than simply one molecule of EFIα interacting with DNA. It is of course possible that EFIβ is not multimerizing, but is perhaps associating with a third, as yet unidentified, factor, EFIγ. However, we have chosen to represent the simplest explanation for the gel shift pattern produced in the absence of EFIα. Also indicative of a multisubunit DNA binding complex in the Q-Sepharose fraction minus EFIα, is the observation that when EFIα (the bound fraction from a 50% glycerol column) is brought to 5% glycerol and refractionated on Q-Sepharose, the EFI DNA binding activity recovered in the bound fraction can be increased by subsequent dialysis to 50% glycerol (data not shown). This may be due to the destabilization of the multimeric homologous (EFIα), complex by low glycerol which can be partially restored, but with very low efficiency, when once again exposed to high glycerol (Fig. 3).

### Isolation of a cDNA Encoding a Polypeptide with Properties of EFIα

Since it appears that EFIα independently possesses specific DNA binding activity, it should be possible to obtain a cDNA clone encoding this polypeptide by screening an expression library with a high affinity binding site for EFI. We have shown previously that EFI binding activity is ubiquitous and may be identical or highly related to CBF and NF-Y (Faber and Sealy, 1990). An available rat liver cDNA expression library (Mueckler and Pitot, 1985), which might be expected to contain clones for EFI/CBF, was screened for plaques capable of binding to a ligated 41-bp oligonucleotide containing the sequence from −121 to −155 in the RSV LTR which binds EFI (and CBF) with high affinity (Faber and Sealy, 1990). The general strategy used was a modification of that described by Landschulz et al. (1988a). One positive clone out of 1.5 × 10⁶ plaques was identified and purified. Initial experiments (data not shown) using the purified phage revealed a general pattern of DNA binding specificity expected for a factor with the properties of EFI. These preliminary observations encouraged us to pursue the further characterization of the cDNA by subcloning these sequences into an inducible plasmid-based expression vector in bacterial cells.

**Bacterially Expressed EFIα Exhibits Specific DNA Binding**—The cDNA purified as described above was cloned into the pBT expression vector system devised by Studier and coworkers (Studier et al., 1990). This system places the protein product of interest under the control of the T7 α10 promoter and results in the expression of a fusion protein linked to the

<table>
<thead>
<tr>
<th>Lanes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5% glycerol flow through</td>
</tr>
<tr>
<td>2</td>
<td>5% glycerol flow through, mixed</td>
</tr>
<tr>
<td>3</td>
<td>5% glycerol column-bound</td>
</tr>
<tr>
<td>4</td>
<td>5% glycerol column-bound, mixed</td>
</tr>
<tr>
<td>5</td>
<td>50% glycerol flow through</td>
</tr>
<tr>
<td>6</td>
<td>50% glycerol flow through, mixed</td>
</tr>
<tr>
<td>7</td>
<td>50% glycerol column-bound</td>
</tr>
<tr>
<td>8</td>
<td>50% glycerol column-bound, mixed</td>
</tr>
<tr>
<td>9</td>
<td>50% glycerol column-bound, mixed, 5% glycerol recovery</td>
</tr>
</tbody>
</table>

![Figure 2: Heterologous and homologous forms of EFI DNA binding activity. Lanes 1-4, the flow-through from a Q-Sepharose fractionation of chick embryo nuclear extract in 50% glycerol shown in Fig. 1, lane 2, was diluted in buffer Q to 5% glycerol and reapplied to Q-Sepharose resin equilibrated in buffer Q containing 5% glycerol. The column flow-through and a high salt elution were collected as described under “Materials and Methods” employing buffers containing 5% glycerol. The column flow-through and a high salt elution were collected, as described under “Materials and Methods,” employing buffers containing 5% glycerol. The 50% glycerol Q-Sepharose flow-through (lane 1) and the 5% glycerol Q-Sepharose column fractions were assayed either individually (lanes 2 and 3) or mixed (lane 4) for EFI DNA binding activity with the radiolabeled 41-bp EFI oligonucleotide. Lanes 5-9, the column-bound material from a Q-Sepharose fractionation of chick embryo nuclear extract in 50% glycerol shown in Fig. 1, lane 3, was diluted in buffer Q to 5% glycerol and reapplied to Q-Sepharose resin equilibrated in buffer Q containing 5% glycerol. The column flow-through and a high salt elution were collected, as described under “Materials and Methods,” employing buffers containing 5% glycerol. The 50% glycerol Q-Sepharose bound material (lane 5) and the 5% glycerol Q-Sepharose column fractions were assayed either individually (lanes 6 and 7) or mixed (lane 8) for EFI DNA binding activity with the radiolabeled 41-bp EFI oligonucleotide. In lane 9, the column bound material shown in lane 7 was also assayed for EFI DNA binding activity after mixing with the Q-Sepharose flow-through from Fig. 1, lane 6. The final glycerol concentration in all gel mobility assay samples was 5%. 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. |
first 11 amino acids of the gene 10 protein. T7 RNA pol-

merase is expressed in the host bacterial cells under the control of the lacUV5 promoter, which is inducible by IPTG. Soluble

extracts were prepared from bacteria carrying either the pET3b vector (which contains no insert) or the pET3b+EFlA construct which contains the subcloned EFIa cDNA. Incubation of these extracts with a 41-bp EFI DNA oligonucleotide results in the formation of a poly(dI):poly(dC)-resistant complex which is present only in induced cells which contain the EFI cDNA insert (Fig. 4A, lane 8). This complex is competed by excess unlabeled wild type EFI 41-bp oligonucleotide (Fig. 4B, lanes 3–5), but not by the mutant oligonucleotide, L59 (Faber and Sealy, 1990) which lacks the inverted CCAAT motif (Fig. 4B, lanes 6–8). A point mutant in the inverted CCAAT motif, which carries a G to T transversion at -129 (ACAAT), and does not bind EFI in nuclear extracts (Faber and Sealy, 1990) also fails to compete for the bacterial EFI complex (Fig. 4B, lanes 9–11). We conclude that the bacterially expressed protein product encoded by the cDNA insert selectively binds to DNA containing the CCAAT motif. We note that the migration of the bacterial protein DNA complex is not that expected for the heterologous (EFIA)(EFIo) or multisubunit EFI complexes identified in nuclear extracts of avian cells (compare Fig. 4A, lanes 8 and 9). The complex obtained with the induced bacterial extracts migrates more rapidly than the eukaryotic counterpart implying a simpler organization which nonetheless possesses specific DNA binding activity.

Bacterially Expressed EFIa Has the Same Molecular Weight as Native EFIa—Extracts obtained from IPTG-induced bacterial cells harboring either the pET3b or pET3b+EFlA constructs were separated on an SDS-polyacrylamide gel and transferred to Immobilon paper. After cutting each lane into a series of slices, the proteins were eluted from each slice and renatured by the approach previously described (Faber and Sealy, 1990). The binding activity of the various eluted fractions as a function of molecular weight is shown in Fig. 5. The peak of EFI DNA binding activity is found at a molecular weight of 35,000. Such binding activity is detected only in the induced bacteria containing the EFIa cDNA. This molecular weight is consistent with that expected from the size of the open reading frame in the inserted cDNA (see below) and is also similar to that reported for the bona fide factor A from eukaryotic cells (Celada and Maki, 1989; Hatamochi et al., 1988).

Bacterially Expressed EFIa Binds Q-Sepharose, even though the Protein Is Overall Positively Charged—The
CCAAT-binding factor EFI_A is characterized biochemically by its ability to bind to a Q-Sepharose resin (Faber and Sealy, 1990). Accordingly, we have assayed if the bacterially expressed protein is capable of binding to Q-Sepharose. Extracts were prepared from IPTG-induced bacterial cells containing vector alone or containing the EFI cDNA insert. Fig. 6 shows the binding of the unfractionated extracts to the EFI DNA (lanes 3 and 8). As seen previously, the induced cells carrying the cDNA insert show the formation of a protein DNA complex of characteristic mobility. The extracts were fractionated on Q-Sepharose following the protocol described previously. Fig. 6, lanes 4, 5, 9, and 10 show the gel shift behavior of material which does not bind to the resin. The specific bacterial EFI protein DNA complex is absent, indicating that the protein product of the EFI cDNA insert is firmly interacting with the resin. Elevation of the ionic strength to 1 M NaCl leads to the elution of the specific bacterial EFI binding activity from the resin (Fig. 6, lane 11).

Thus, the EFI cDNA clone we have isolated encodes a polypeptide which requires the CCAAT motif for specific DNA binding and binds to the anionic Q-Sepharose resin, two properties which are consistent with its identity as EFI_A. Interestingly, inspection of the amino acid composition of the open reading frame showed contained within the EFI_A cDNA indicates that the protein has an overall positive charge, yet it is able to bind to the positively charged chromatography resin. Similar behavior has been reported for other DNA binding factors (Shasta et al., 1982). In the case of EFI_A, it most probably reflects the clustering of negative charges arising from the unusual distribution of charge throughout the molecule (see below).

**Sequence Analysis of the EFI_A cDNA Clone Reveals Near Identity with the Human CCAAT Binding Factor, YB-1**—The EFI_A cDNA, isolated and purified from the λgt11 vector, was subcloned into pGEM-7zf(+) and sequenced by the didexy method. The 1489-nucleotide EFI_A cDNA sequence is shown in Fig. 7. The ATG sequence is shown at position 100 as the translation start site or very near to the translation start site. This is based on the Northern analysis of the mRNA of rat hepatocyte H4 cells shown in Fig. 8, which indicates a single discrete 1600-nucleotide species hybridizing to the EFI_A cDNA sequence. Clearly, this indicates that most of the information in the EFI_A mRNA has been retained in the ~1500-bp cDNA. This leaves only 100 or so bases which might contain an additional translational start. Use of the ATG at position 100 as the initiation codon gives rise to a 322-amino acid open reading frame, the sequence of which is presented in Fig. 9. 343 and 399 nucleotides beyond the...
The nucleotide sequence of the 1489-bp EFI, cDNA determined by the human YB-1 cDNA from Didier et al. (1988) is given on the bottom line. Lines are used to denote positions of exact match. Given on the top line. A comparison with the nucleotide sequence of Locations of the open reading frames encoding the EFIA and YB-1 FIG. 7. Nucleotide sequence of the EFIA and YB-1 cDNAs.

Locations of the open reading frames encoding the EFIA and YB-1 proteins are indicated by arrows. The nucleotide sequences were aligned using the University of Wisconsin Genetics Computer Group BESTFIT software program. The amino acid sequences of EFIA, YB-1, and dbpB are nearly identical. The deduced amino acid sequences of EFIA, YB-1, and dbpB, obtained from the nucleotide sequence presented in Fig. 7, from Didier et al. (1989), and from Sakura et al. (1988), respectively, are compared. Lines are used to denote positions of exact match. The amino acid sequences are aligned using the University of Wisconsin Genetics Computer Group BESTFIT software program. Termination codon we find two polyadenylation sites at positions 1411 and 1467 bp. The sequence of the EFIA cDNA shows extensive homology (93%) with that reported previously for YB-1 (Didier et al., 1988), a CCAAT binding factor which recognizes that Y-box in the promoter of the human HLA-DR (Y chain gene (nucleotide and amino acid sequences from Didier et al. (1990), and from Sakura et al. (1988), respectively, are compared. Lines are used to denote positions of exact match. The amino acid sequences are aligned using the University of Wisconsin Genetics Computer Group BESTFIT software program.

FIG. 8. Size determination of the rat EFIA mRNA. Total RNA from the rat H4IIE hepatoma cell line was isolated and analyzed by gel electrophoresis. Upon transfer to a nitrocellulose filter, the RNA was probed with a randomly primed EFIA cDNA probe. Autoradiography of the filter is shown; positions of migration of the 18 and 28 S RNAs determined by ethidium staining of the RNA sample in the gel are indicated.

FIG. 9. Amino acid sequences of EFIA, YB-1, and dbpB are nearly identical. The deduced amino acid sequences of EFIA, YB-1, and dbpB, obtained from the nucleotide sequence presented in Fig. 7, from Didier et al. (1989), and from Sakura et al. (1988), respectively, are compared. Lines are used to denote positions of exact match. The amino acid sequences are aligned using the University of Wisconsin Genetics Computer Group BESTFIT software program.

termination codon we find two polyadenylation sites at positions 1411 and 1467 bp. The sequence of the EFIA cDNA shows extensive homology (93%) with that reported previously for YB-1 (Didier et al., 1988), a CCAAT binding factor which recognizes that Y-box in the promoter of the human HLA-DR (Y chain gene (nucleotide and amino acid sequences from Didier et al. (1990), and from Sakura et al. (1988), respectively, are compared. Lines are used to denote positions of exact match. The amino acid sequences are aligned using the University of Wisconsin Genetics Computer Group BESTFIT software program.
of the YB-1 protein deviate from the EFIα sequence, and the rat protein contains an additional seven amino acids at the carboxyl terminus. Other than these changes we find that the amino acid sequences are identical, resulting in an overall homology between the two proteins of 95%. Interestingly, the homology between the two cDNAs encompasses not only the coding region, but also extends into the 5'- and the 3'-untranslated regions (see Fig. 7). The positions of two canonical poly(A) signals are at similar positions in both molecules.

In addition, we also find extensive homology between the cDNAs of EFIα and dbpB, a DNA binding protein of unknown sequence specificity whose cDNA was obtained by screening a human placental cDNA expression library with DNA probes containing the EGF receptor enhancer or the promoter of the c-erbB-2 gene (Sakura et al., 1988). A second cDNA clone, dbpA, was also obtained by this approach (Sakura et al., 1988).

It was recently noted that the E. coli cold shock protein, CS7.4, and three human proteins, dbpA, dbpB, and YB-1, share a region of conserved sequence encompassing 72 amino acids (located from position 55 to 127 in YB-1) postulated to represent a DNA binding motif (Wistow, 1990). In fact, the amino acid sequence homology between YB-1 and dbpB extends throughout their entire 318 or 324 residues, respectively. The deduced amino acid sequence of dbpB has also been presented in Fig. 9 for comparative purposes. It is 97% homologous with EFIα and shares the same carboxyl terminus, including the additional seven amino acid residues at the carboxyl terminus which YB-1 lacks. Other than this difference at the carboxyl terminus, dbpB and YB-1 are identical except for an alanine/glutamic acid substitution at position 120, where dbpB (alanine 120) corresponds to EFIα. Given the near identity with YB-1 and EFIα, it is likely that dbpB is also a CCAAT-binding protein.

Finally, we find significant homology between EFIα and two other factors known to be CCAAT binding proteins, FRG Y1 and FRG Y2. FRG Y1 and FRG Y2 were identified and cloned from Xenopus laevis by Tafuri and Wolffe (1990). The protein sequence homology between FRG Y1 and EFIα amounts to 263/322 amino acids with absolute identity between EFIα residues 46-162 and FRG Y1 residues 28-142. The homology between FRG Y2 and EFIα is only 184/322; however, there is almost perfect identity between EFIα residues 50-140 and FRG Y2 residues 35-125.

A Novel Structural Motif in the EFIα Polypeptide—Analysis of the protein sequence derived from the nucleotide sequence for EFIα indicates that this protein is, in general, highly hydrophilic, with an unusual number of charged residues. Overall the molecule is positively charged (net +17). Inspection of the charged residues reveals a remarkable distribution. In general, residues with like charges are grouped together, so that alternating domains of positive and negative charge are formed as shown in Fig. 10. This type of organization is noted from residue 90 to near the carboxy terminus. When the molecule is viewed in this light it becomes apparent that other residues are distributing themselves in a regular manner among the repeating charge domains. Thus the positive regions are characterized by sequences rich in arginine, tyrosine, and proline; the negative regions contain primarily, although not exclusively, glutamic acid along with elevated levels of glycine. The short spacer regions between the alternating charge domains are enriched in glutamine. The importance of the alternating regions and the intervening glutamine-rich regions is supported by their maintained appearance in the frog genes FRG Y1 and FRG Y2, even though surrounding amino acid residues have been permitted to change. Clusters of alternating charged residues are also found in dbpA, although the enrichment of glutamine residues in the intervening regions is not seen.

DISCUSSION

A number of CCAAT binding proteins, including EFIα, have been demonstrated to be composed of two heterologous polypeptides, both of which appear to be required for specific DNA binding activity (Celada and Maki, 1989; Chodosh et al., 1989; Hatamochi et al., 1988; Olesen et al., 1987). This conclusion is based primarily on mixing experiments performed after the two components had been separated via ion exchange chromatography. We have identified conditions (high glycerol) under which EFIα DNA binding activity is not separated during ion exchange chromatography. In fact, we have found that two forms of EFI DNA binding activity, with different ion exchange properties, exist; the heterologous (EFIα)(EFIβ) form, and a form lacking EFIβ, which in the simplest scenario may be a multimeric EFIα complex. We have recently found that EDTA can also influence multimer stability (Faber and Sealy, 1989); high levels of EDTA stabilize both heterologous and homologous multimers, whereas low levels (0.1 mM) promote dissociation during Q-Sepharose chromatography. Once dissociated, EFIβ can be readily recombined with EFIα to regenerate the heterologous form of EFI DNA binding activity in vitro, whereas association of EFIβ to regenerate the homologous form of high affinity EFI DNA binding activity in vitro is very inefficient. We do not know whether the heterologous and homologous forms of EFI DNA binding activity also occur in vivo or whether either might represent an artifact of the nuclear extraction procedure. Nonetheless, our finding that EFIα possesses high affinity specific DNA binding activity independent of EFIβ prompted us to screen an expression library with radiolabeled EFI DNA sequences to determine if a cDNA clone for the EFIα polypeptide could be identified in this manner.

Isolation of a cDNA Clone Encoding a Polypeptide with Similarity to EFIα—Screening of a rat liver cDNA expression library with radiolabeled EFI DNA sequences yielded one...
positive clone. The bacterially expressed product from the cDNA insert in this clone showed properties typical of those expected for EFI+. Thus it is a DNA binding protein which binds selectively to sequences containing the intact inverted CCAAT motif. The molecular weight of the bacterial EFI+ (35,000) is comparable with that reported for identical or close homologs of this factor from mammalian cells CBFα (58,000; Hatamoto et al., 1988) and Y-box factor A (34,000; Celada and Maki, 1989). As yet, we have not completed genomic cloning and primer extension analysis to determine the precise start of translation. It is entirely possible that the ~35,000 dalton product we are describing may not contain the full sequence of the native eukaryotic protein, and indeed this possibility is reflected in an earlier report from this laboratory (Faber and Sealy, 1990) on the molecular weight of avian EFI+, (~43,000–60,000). Nonetheless, such a possible truncation does not appear to have had a pronounced effect on the specificity of the interaction of this protein with DNA. The bacterially expressed protein binds to the cationic resin Q Sepharose. This property, although consistent with that of its eukaryotic counterpart, is somewhat surprising given its overall positive charge. However, the charged amino acid residues in this protein are organized into discrete regions, and this may permit the protein to develop a three-dimensional domain of negative charge necessary for binding to the ion exchange resin. Thus, by most of the criteria defined by the eukaryotic protein we can assert that the bacterial protein is made from a cDNA which appears to be that for rat EFI+. However, we note that one activity of the eukaryotic EFI+, which has not as yet been duplicated by the bacterially expressed EFI+, is the ability to form multimeric complexes with either EFI+ or EFIα (bacterially expressed or avian) leading to the gel shift pattern seen with EFIα in avian nuclear extracts. This may reflect a lack of post translational modifications or possibly might be due to the presence of a truncated form of EFI+, which is unable to complex with additional avian EFIα or EFI, polypeptides in the usual manner.

**EXTENSIVE HOMOLOGY BETWEEN THE CDNAS ENCODING EFI+ AND OTHER CCAAT BINDING PROTEINS**—The nucleotide sequence of the rat EFI+ cDNA shows extensive homology with two previously reported human cDNAs encoding dbpB (Sakura et al., 1988) and another CCAAT binding protein, YB-1 (Didier et al., 1988). The DNA binding specificity of dbpB is unknown. However, the amino acid sequence within the open reading frames of these three cDNAs are nearly identical, rendering it likely that dbpB is also a CCAAT binding protein. Indeed, a CCAAT motif is located at approximately -80 in the c-erbB-2 gene promoter (Ishii et al., 1987), and an inverted CCAAT motif also occurs in the 530-bp epidermal growth factor downstream enhancer fragment (Maekawa et al., 1989) used to identify dbpB (and dbpA).

We have argued previously that EFIα may be identical or closely related to two other CCAAT binding proteins, CBF and NF-Y (a factor which binds to the Y-box of the murine MHC class II Eα gene). This was based upon the equal or greater affinity of EFIα in chick embryo nuclear extracts for CBF and NF-Y binding sites, relative to the EFI site in the RSV LTR, as well as similar biochemical fractionation behavior (with CBF) and methylation interference patterns (with NF-Y) (Faber and Sealy, 1990). The finding that the EFIα cDNA encodes a polypeptide nearly identical to a human Y-box binding protein lends additional support to this conclusion. YB-1 and NF-Y have not been shown to be two-component CCAAT binding proteins, although another Y-box binding protein which recognizes the Y-box of the murine MHC class II gene I-Aβ fractionates identically to EFIα on a Mono Q resin in low glycerol (Celada and Maki, 1989). However, we have demonstrated here that buffer conditions greatly influence whether or not specific DNA binding is dependent on mixing of two heterologous components after ion exchange fractionation. It is interesting that the extensive homology between the EFIα and YB-1 or dbpB cDNAs extends beyond the coding region into both the 5'- and 3'-untranslated regions. This suggests that important functional elements may be located in these regions as they have apparently not been allowed to diverge. For example, analysis of the nucleotide sequence of EFI+ cDNA reveals a nearly perfect 30-bp direct repeat immediately upstream of the presumed translational start site (position 100) as shown below.

```
4CGCCGCCCGGGCCCGGTCAGACAC
5CCTCGCGCGCGGGGCCAGCTCACCAC
```

Inspection of this region in Fig. 7 indicates that very similar repeats are found in the human YB-1 (and dbpB) cDNAs (and also in the frog FRG Y1 and Y2 cDNAs). It seems unlikely that such motifs should have occurred and been preserved by chance, and this may be an indication of protein binding domains in either the RNA or the chromosomal DNA.

In addition to the near identity with YB-1 and dbpB, the EFIα cDNA shares extensive homology with two X. laevis cDNAs (FRG Y1 and FRG Y2) encoding CCAAT binding proteins. A region comprised of EFIα residues 50–140 is basically invariant in all five CCAAT factors. Moreover, the corresponding region in YB-1 and dbpB (from amino acid 52, extending to amino acid 156) is identical with the central region of dbpA (Sakura et al., 1988; Wisotz, 1990). This region of common amino acid sequence would therefore appear to be a likely candidate for that portion of the molecule involved in binding to the CCAAT recognition motif in the DNA. If so, it suggests that dbpA is also a CCAAT binding protein, although dbpA diverges considerably from EFIα/YB-1/dbpB and FRG Y1 or Y2 outside the highly conserved domain. Within this conserved domain (60–140/154, numbering with respect to EFIα here and below), no strong homologies to the zinc finger or helix-loop-helix DNA binding motifs exist. However, there is homology between residues 115 and 123 and the helix-turn-helix motif of the lαCI transcriptional activator protein (Ito et al., 1998) limited to the amino-terminal helix. It is interesting that positions 53–125 within this conserved domain closely match the entire deduced amino acid sequence of the cold-shock protein, CS7.4 (Wisotz, 1990). The significance of this homology to a putative CCAAT recognition motif for the function of CS7.4 is unknown.

Finally, we have noted a novel arrangement of charged residues in the EFIα coding region. Alternating clusters of positively and negatively charged residues occur throughout the carboxy-terminal two thirds of the molecule. The positive clusters in this CCAAT binding polypeptide are rich in proline, tyrosine, and groups of arginine residues (PYRR domains), whereas the negative regions contain primarily glutamic rather than aspartic acid and are also rich in glycine. The spacer regions between the charged domains are enriched in glutamine. The significance of this striking organization of charged residues is supported by their maintenance in the frog CCAAT binding proteins and also in dbpA (although glutamine enrichment in the intervening regions is not seen in dbpA). Several different functional roles can be envisioned for this alternating arrangement of charge clusters. Since EFIα most likely functions as a ubiquitous transcription factor in the form of a multihomotetrameric species, it is possible that the function of the alternating charge regions in the carboxy tail...
may be to act as a charge "zipper" facilitating the formation of homomeric, (EFIA)$_n$ and/or heteromeric, (EFIA)(EFIB), species. Alternatively, the alternating charged residues may interact intramolecularly to form the intervening glutamine-rich regions into a series of glutamine "fingers" to generate, for example, a surface necessary for transcriptional activation. Site-directed mutagenesis and domain-swapping experiments are now in order to experimentally define the functional domains within the EFIA polypeptide responsible for specific DNA binding, protein-protein interactions, and transcriptional activation.

Acknowledgments—We thank our colleagues in the Scaly and Chalkley laboratories for their support, and we are indebted to George Amorino, Karen Marcus, and Kathy Mattix for excellent technical assistance. We thank the B. Hogan laboratory for generously providing the pETBb, BLPl(DE3)pLysS host/vector expression system and T. Weil for carefully reading the correspondence in Nature. We owe a particular debt of gratitude to Dr. Alan Wolffe for helpful discussions and generous sharing of sequence information prior to publication.

REFERENCES

Dyman, W. S. (1989) Cell 58, 1–4