Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

Activity and tissue-specific expression of the transcription factor NF-E1 multigene family

Masayuki Yamamoto, Linda J. Ko, Mark W. Leonard, Hartmut Beug,[1] Stuart H. Orkin,[2] and James Douglas Engel[3]

Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, Illinois 60208-3500; [1]Institute of Molecular Pathology, A1030 Vienna, Austria; [2]Division of Hematology/Oncology, Children's Hospital and Department of Pediatrics, Harvard Medical School, Howard Hughes Medical Institute, Boston, Massachusetts 02115

NF-E1, a DNA-binding protein that recognizes the general consensus motif WGATAR, is the first tissue-specific factor to be identified in erythroid cells. Using a probe from the murine GF-1 (NF-E1) cDNA clone, we isolated three homologous chicken cDNAs: One of these corresponds to an mRNA (NF-E1a) that is abundantly and exclusively expressed in erythroid cells; a second mRNA (NF-E1b) is also expressed in all developmental stages of erythroid cells but is additionally found in a limited subset of other chicken tissues; mRNA representative of a third gene (NF-E1c) is expressed only in definitive (adult) erythrocytes within the red cell lineage but is also abundantly expressed in T lymphocytes and brain. All NF-E1 proteins are highly conserved within the DNA-binding domain and bind to the consensus motif with similar affinities in vitro; they are also all stimulatory *trans*-acting factors in vivo. The factors differ quantitatively in their ability to *trans*-activate reporter genes in which the number and position of cognate binding sites is varied relative to the transcriptional initiation site. These data suggest that the NF-E1 consensus motif directs a broader and more complicated array of developmental transcriptional regulatory processes than has been assumed and that NF-E1c may play a unique regulatory role in the developing chicken brain and in T lymphocytes.

[*Key Words*: Transcription factors; erythroid cell; T cell; *trans*-activation]

Received May 30, 1990, revised version accepted July 17, 1990.

Our current view of eukaryotic gene transcriptional regulation is based primarily on the observation that a wide variety of *cis*-acting DNA sequences are able, both singly and in concert, to aid or inhibit productive transcriptional initiation. These regulatory interactions have now been shown in many cases to be affected by the differential, site-specific DNA binding of constitutive and tissue-specific regulatory *trans*-acting factors that act to either enhance or inhibit the ability of RNA polymerase and ancillary transcription factors to form a functional initiation complex (for review, see Maniatis et al. 1987; Ptashne 1988; Johnson and McKnight 1989). Furthermore, *cis*-regulatory domains appear to exert their effect by binding various combinations of both constitutive and tissue-specific factors to form functional modules; however, only rarely does the effect of deletion of a single protein binding site within a module abrogate the overall physiological effect of the entire regulatory module (e.g., Gallarda et al. 1989).

β-Globin gene expression provides a well-defined system for analysis of the developmental, genetic, and biochemical basis of transcriptional regulation. β-Globin expression is regulated primarily at the level of transcriptional initiation (Groudine et al. 1981). Molecular genetic analysis has implicated the presence of several levels of transcriptional control: The first appears to be the ability to form an active chromatin domain (Weintraub and Groudine 1976) to allow the β-globin gene cluster to be transcribed; such activity may be a reflection of the genetic effect of dominant control regions (DCRs) or locus activation regions (LARs) within the β-globin gene cluster (Forrester et al. 1987; Grosveld et al. 1987; Trudel et al. 1987). Another level of control appears to be elicited by the tissue-specific transcriptional activity of the β-globin gene enhancer (Choi and Engel 1986; Hesse et al. 1986; Behringer et al. 1987; Kollias et al. 1987), a complex module whose activity is regulated by the binding of both ubiquitous and tissue-specific factors (Emerson et al. 1987; Wall et al. 1988; Gallarda et al. 1989). The final level of regulation, determination of which globin gene isotype is to be expressed at a particular developmental stage (referred to as hemoglobin switching), is elicited by cooperative *cis*-regulatory interactions between distal control elements (enhancers or DCR/LAR) and sequences within physically linked genes (Choi and Engel 1988; Enver et al. 1990). In

[3]Corresponding author.

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

at least one case, this regulation may be due to the binding, within the chicken adult β-globin gene promoter and enhancer, of a unique developmental stage- and tissue-specific erythroid transcription factor, NF-E4 (Gallarda et al. 1989).

Within the chicken β-globin gene promoter and enhancer, multiple constitutive and tissue- or stage-specific *trans*-acting factor binding sites have been defined, and their activity has been correlated with function both in vitro and in vivo (Emerson et al. 1987, 1989; Choi and Engel 1988; Lewis et al. 1988; Nickol and Felsenfeld 1988; Reitman and Felsenfeld 1988; Gallarda et al. 1989). The first of the erythrocyte-specific *trans*-acting factor binding sites to be identified is a prevalent motif, WGATAR (W = T or A; R = G or A), found in common within the human γ-globin promoter and in the human and chicken β-globin gene enhancers (Evans et al. 1988; Wall et al. 1988; Catala et al. 1989; Martin et al. 1989; Perkins et al. 1989). This element has since been found to be associated with a variety of erythroid gene regulatory regions (Trainor et al. 1987; Knezetic and Felsenfeld 1989; Mignotte et al. 1989; Plumb et al. 1989). These independent discoveries also assigned different names to the factor(s) recognizing the consensus motif, variously denoted NF-E1, GF-1, Eryf1, and EF-1.

The initial cDNA cloning and analysis of murine, chicken, and human NF-E1 (Evans and Felsenfeld 1989; Tsai et al. 1989; Trainor et al. 1990; Zon et al. 1990) demonstrated that the proteins produced by this erythroid-specific mRNA are able to recognize the consensus DNA binding site in vitro and that the mRNAs encoding these factors are transcribed only in erythroid cells. However, it is not clear from such studies whether or not the isolated chick, mouse, and human cDNA clones are functional homologs of one another or

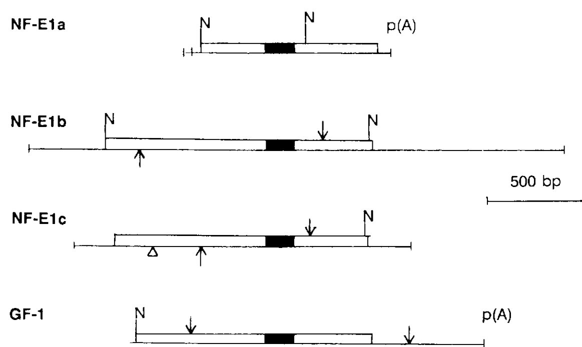


Figure 1. Structure of NF-E1 cDNAs. cDNA clones with a high degree of identity to the finger region of murine NF-E1 (GF-1; Tsai et al. 1989) were isolated from chicken reticulocyte and embryo cDNA libraries and physically characterized by restriction enzyme mapping and DNA sequence analysis (see Results). Abbreviated restriction enzyme maps of the three different loci isolated (NF-E1a, NF-E1b, and NF-E1c) are shown. The chicken and murine proteins are aligned with respect to the finger domains (solid box). The open box represents the extent of each of the predicted ORFs within the cDNA clones; the solid line represents the extent of cloned cDNA segments (Figs. 2 and 3). (N) *Nco*I; (↓) *Eco*RI; (↑) *Bam*HI; (△) *Kpn*I; [p(A)] polyadenylation consensus sequence.

whether the proteins are capable of exerting identical activities in vivo. Thus, a functional analysis of chicken NF-E1 was initiated, in which we have compared this activity to the murine factor GF-1.

We used the murine GF-1 clone (Tsai et al. 1989) to isolate the chicken homolog of this factor. To our surprise, three distinct NF-E1 cDNAs were found to represent a multigene family, each of whose members is highly tissue restricted in expression. We find that all three chicken NF-E1 proteins bind with high affinity to the consensus recognition sequence in vitro and also that the chicken NF-E1 proteins are able to stimulate transcription from reporter genes in vivo, but differ in ability to stimulate transcription from reporter constructs bearing different numbers of factor binding sites or sites located in different positions relative to the start of transcription. These studies suggest that the different NF-E1 *trans*-acting factors, containing indistinguishable *cis*-regulatory DNA sequence recognition properties, differentially contribute to the program of developmentally regulated transcriptional activation in a variety of chicken tissues.

Results

Isolation and sequence analysis of chicken GF-1 homologs

A DNA fragment corresponding to the DNA binding domain of the murine GF-1 cDNA clone (Tsai et al. 1989) was used as an initial probe to screen a cDNA library prepared from anemic hen reticulocyte mRNA (RBC4; see Methods), whereas probes derived from these clones were used in further screening of a second library prepared from whole 10-day-old chicken embryos (Sap et al. 1986). Ninety-one independent recombinants were isolated, representing three separate genomic loci, designated NF-E1a, NF-E1b, and NF-E1c (Fig. 1).

Double-stranded DNA sequencing was performed on the largest cDNAs representing each locus. The NF-E1a DNA sequence (data not shown) is identical, within the coding region, to the published sequence of Eryf1 (Evans and Felsenfeld 1989). The sequence predicts a 304-amino-acid NF-E1a protein of 31,415 daltons (pI 10.2). The NF-E1b cDNA (p30a) is 2812 bp in length and contains a single long open reading frame (ORF) encoding 466 amino acid residues predicting a protein of 50,147 daltons (pI 9.9). The methionine codon at the 5′ end of this ORF is located at nucleotide 410: The sequence surrounding the codon fits the ideal consensus for translation initiation (Kozak 1989; Fig. 2). The NF-E1c cDNA clone (p31a) predicts a single ORF from nucleotide position 187 to 1518 (Fig. 3). Conceptual translation of this ORF gives a predicted size for the NF-E1c protein of 48,194 daltons (pI 9.9), and once again, the sequence surrounding the 5′-most methionine codon is in a good translation context. An in-frame termination codon directly precedes the assigned ATG initiation codon of NF-E1c (nucleotide 4, Fig. 3) providing evidence that the ORF encodes the genuine NF-E1c protein.

When aligned for maximum identity to GF-1, the se-

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press


```
                                    CCCGCGCCGGGGGTCTCTCCGACCCTCCATCGGAAAATCTCCCCGCGCCCCGCGGCGATCCGGCGCGGCCCCCACT    76
GCCGCTCTCGGCGCTGTCCCCGCGGCCAACCCCGGCGCCCGCCGACCCCATGCAGCCTCCGGGGCTCGGCGAGGGACCCCGTCCCCGCACACACGCACGCACCGGGAGGA   187
AAAGGCACAGCAGGTGCTGCCCCGCCACCCGCCCGCGGACAGAGGGACGCGCTGACTCCGGGGCCCACCCCGCATCCTCCCGAGGGCGATCTCCGGACACCCGACGGCTTT   298
CTCTTAAACAATCGCCCGTCTGCTAACCCCCCCCTCCGGTGCCTCCGCCTGCCGACCCCGACCGCCGCTCGCCGCTGCCCCGCGGGTGGGCGGAGGGAGGCCCGAGGCC   409

ATG GAG GTG GCC ACG GAT CAG CCT CGC TGG ATG ACC CAC CAC GCC GTG CTC AAC GGG CAG CAC CCC GAG AGC CAC CAC CCG GGA   493
 M   E   V   A   T   D   Q   P   R   W   M   T   H   H   A   V   L   N   G   Q   H   P   E   S   H   H   P   G
CTG GCT CAC AAC TAC ATG GAA CCA GCG CAG CTC CTA CCT CCG GAC GAA GTC GAC GTC TTC TTC AAC CAC CTG GAT TCC CAG GGC   577
 L   A   H   N   Y   M   E   P   A   Q   L   L   P   P   D   E   V   D   V   F   F   N   H   L   D   S   Q   G
AAC CCT TAC TAT GCC AAC TCT GCC CAT GCC CGG GCC CGG GTC TCC TAC AGT CAG GCA CAC GCC CGC CTG ACC GGG AGT CAG ATG   661
 N   P   Y   Y   A   N   S   A   H   A   R   A   R   V   S   Y   S   Q   A   H   A   R   L   T   G   S   Q   M
TGC CGG CCT CAT CTT ATC CAC AGC CCC GGG ATC CCT TGG CTG GAC AGC AGC AAG GCG GCA CTG TCT GCC CAT CAC CAC AAC CCC   745
 C   R   P   H   L   I   H   S   P   G   I   P   W   L   D   S   S   K   A   A   L   S   A   H   H   H   N   P
TGG ACC GTC AAC CCC TTC ACC AAG ACC CCC CTG CAC CCC TCG GCG GCC GGA GCA CCT GGG GCC ATC TCG GTG TAC CCT GGC AGC   829
 W   T   V   N   P   F   T   K   T   P   L   H   P   S   A   A   G   A   P   G   A   I   S   V   Y   P   G   S
AGC ACG TCC AGC ACC GCC TCC GTC TCG TCG CTC ACC CCC GCC TCG CAC TCG GGC TCC CAC CTC TTC GGT TTC CCC CCG ACC CCT   913
 S   T   S   S   T   A   S   V   S   S   L   T   P   A   S   H   S   G   S   H   L   F   G   F   P   P   T   P
CCC AAG GAA GTG TCT CCA GAC CCC AAC TCC ACC AGC GCT GCC TCC CCT TCG TCC TCC GCT GGG GCC CGG CAG GAG GAC AAA GAC   997
 P   K   E   V   S   P   D   P   N   S   T   S   A   A   S   P   S   S   S   A   G   A   R   Q   E   D   K   D
AGC ATC AAG TAC CAA GTG TCG CTG TCG GAA GGG ATG AAG ATG GAG AGC GCC AGC CCG CTC CGC AGC AGC CTC ACC AGC ATG GGG  1081
 S   I   K   Y   Q   V   S   L   S   E   G   M   K   M   E   S   A   S   P   L   R   S   S   L   T   S   M   G
GCC CAG CCC TCC ACC CAC CAC CCC ATC CCC ACA TAC CCC TCT TAC GTG CCG GCT GCC CAT GAC TAC AGC AGC AGC CTC TTC CAC  1165
 A   Q   P   S   T   H   H   P   I   P   T   Y   P   S   Y   V   P   A   A   H   D   Y   S   S   S   L   F   H
CCG GGC AGC TTC CTG GGG GGC CCG GCG TCC AGC TTC ACC CCC AAG CCA CGA AGC AAG GCC AGA TCC TGT TCA GAA GGC AGA GAG  1249
 P   G   S   F   L   G   G   P   A   S   S   F   T   P   K   P   R   S   K   A   R   S   C   S   E   G   R   E
TGT GTG AAC TGT GGA GCA ACC GCT ACC CCT CTC TGG AGA AGA GAC GGC ACC GGG CAT TAC CTG TGT AAC GCC TGC GGG CTC TAC  1333
 C   V   N   C   G   A   T   A   T   P   L   W   R   R   D   G   T   G   H   Y   L   C   N   A   C   G   L   Y
CAC AAA ATG AAC GGT CAA AAC CGA CCT CTC ATT AAA CCC AAA CGA AGG CTG TCA GCG GCC AGG AGA GCA GGG ACG TGT TGT GCC  1417
 H   K   M   N   G   Q   N   R   P   L   I   K   P   K   R   R   L   S   A   A   R   R   A   G   T   C   C   A
AAC TGT CAG ACA ACC ACC ACG ACC TTA TGG CGA CGC AAC GCC AAC GGG GAC CCG GTT TGT AAT GCC TGC GGA CTC TAC TAT AAA  1501
 N   C   Q   T   T   T   T   T   L   W   R   R   N   A   N   G   D   P   V   C   N   A   C   G   L   Y   Y   K
CTG CAC AAT GTG AAC AGG CCT CTG ACC ATG AAA AAG GAA GGA ATT CAG ACC AGG AAT AGG AAG ATG TCC AAC AAA TCA AAG AAA  1585
 L   H   N   V   N   R   P   L   T   M   K   K   E   G   I   Q   T   R   N   R   K   M   S   N   K   S   K   K
AGC AAG AAA GGC TCT GAG TGT TTT GAG GAA CTG TCC AAG TGC ATG CAA GAG AAA TCG TCT CCT TTC AGC GCC GCT GCC CTT GCT  1669
 S   K   K   G   S   E   C   F   E   E   L   S   K   C   M   Q   E   K   S   S   P   F   S   A   A   A   L   A
AGT CAC ATG GCA CCT ATG GGG CAT TTG CCA CCA TTC AGC CAC TCT GGA CAC ATC CTA CCA ACA CCT ACC CCC ATC CAC CCA TCT  1753
 S   H   M   A   P   M   G   H   L   P   P   F   S   H   S   G   H   I   L   P   T   P   T   P   I   H   P   S
TCC AGC ATC TCA TTT GGA CAT CCG CAT CCA TCC AGC ATG GTT ACA GCC ATG GGA TAA AACCTGACGACCAAGAGACCCACCCAGAGATCGAG  1845
 S   S   I   S   F   G   H   P   H   P   S   S   M   V   T   A   M   G   *

AACATGGGGCTTTACCTAAGACAACACCAAGTAAGAAAGTGTTCTGCCAAGAGGAGAATGTAGGGATGGCAAAAAAAGGGGATCTGTGCTCCCGTGTGGTTTAACTCTTCA  1956
CTCTGGACTAAAGCCTTGCAAGTGATGGGTCTTCTGCAGAAATGTGTAGTGGTGCCTGTAATGCACTTTGCCCCCTCAGGTATAAGGAAAAGAAGAACTCACCTGGTCTAT  2067
ATTCAGTTGTCTAGCAGAAAGCAAAAGGTGGCAGAAGCAGAAAGAAGAGGCTGGAACTGGCTTTCCTTTCTCTCTTTGTTATTACTGTGAATAGTGTAAAGAGATATAAGC  2178
AGCCTCCCAGGATGAAATCGGCTCACTGGAAGCCACACATGCTGGATTTCTATTGTGGACTCTGTTTGGGGAGGGGAAAGAAAAAAGAAAAAAAAAAAAAAGGATGTCTTT  2289
ATAAAAAACTTAATTTTTTAAATGAAATCAGACACAAAGTCATATTTATTTTGCTCTTGTTTTCCACAAAAAGCCTCAACCCCTCTGAGTGCATGTTTTTGTGCTGTTGGT  2400
TGCAATGGCTACCCTGTATGAAAATCTGCCTTCGTAAATATGGGGAGGGACAATATTTTGTGGCCCCCCTGGATTTGCTCCATCTATCTTCCATACCAAAATATCCCTTAG  2511
GAAGGGAAGAAAGAGAGGGTGGGCAAAGCAAGAACTGTAGCCGCAGTATGGGATTTCAATGCACCCAAGACGGTGATTATTTTTGACACAATTTCTTCTTTCATGTTAGGG  2622
AAAAGAAGGACTTATTTAAAAAACAACAACAACAAAAAAACCCCACACAACAGCCCCCCTCCCAGCTCCCCTATTATTCCTCAAAGGAGAGGGCTGCATCAGAGTAACTGG  2733
TGGGAAGACGCTCCTGGAAGGCAGTCCTGCACTAGCGAGCGGTGAAATACCAGATCGCTGCAGCTAACGTGTAGCAATT                                  2812
```

Figure 2. Sequence of NF-E1b cDNA. The nucleotide sequence of cDNA clone p30a is shown; numbers at *right* represent nucleotides from the 5′ end of the insert. The translational initiation codon is assigned on the basis of amino acid sequence identity with the initiation site of NF-E1c (Fig. 3). The shaded area corresponds to the finger domain and the sequences encoding it.

quence predictions of the conceptually translated chicken NF-E1 cDNAs display several noteworthy features (Fig. 4). First, all of the chicken proteins are remarkably similar within the entire presumptive DNA-binding domain (the "fingers;" Evans and Felsenfeld 1989; Tsai et al. 1989); the three genes represented here encode protein products that share >90% sequence identity within this region. A similar degree of conservation exists between all of the chicken proteins and murine sequences within this domain. Second, the amino-terminal domains of the NF-E1b and NF-E1c proteins contain substantially greater identity to the murine GF-1 protein than NF-E1a does. Third, the carboxy-terminal domains of NF-E1a and GF-1 have much greater similarity to one another than either have to NF-E1b or NF-E1c. Finally, NF-E1b and NF-E1c clearly share far greater overall amino acid sequence identity with one another than any of the chicken proteins share with the GF-1 protein.

Tissue specificity of NF-E1 expression

To address the question of whether or not the three factors would differ in expression in various cell types and during maturation of erythroid progenitors, full-length probes were prepared to the three NF-E1 cDNAs (Fig. 1) and used in RNA blot analysis. Several interesting and potentially important regulatory features of the expression pattern of these mRNAs can be deduced from these studies.

NF-E1a mRNA is strikingly abundant, is restricted to expression within the erythroid lineage, and is ~1.3 kb in size (Fig. 5A). Although NF-E1a mRNA accumulation is equal in embryonic and adult reticulocytes (the primitive and definitive lineages, respectively), the mRNA is fourfold less abundant in HD6 cells (Beug et al. 1982), representing a stage in differentiation preceding overt β-globin transcription.

NF-E1b is encoded by a 4.3-kb mRNA and is expressed at comparable levels in all stages of erythroid cells, as well as in embryonic brain and liver, adult kidney, embryonic cardiac muscle, and fibroblast (Fig. 5B). NF-E1b is not detectable in a T-lymphocyte cell line, adult liver, or embryonic skeletal muscle. In definitive reticulocytes, NF-E1b mRNA appears to be present at ~10% of the level of erythrocyte band 3 or β-actin mRNAs (see

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press


```
                                  TTTTAAATTTTTGGAACCGATCATTTCTCACCACCACCCCCCTCTTCCCTCCGATTACAAACCAACCTAACCCCA     75
ACCTCGCCGAAGCTCTTTCCCCACCCCGACTCTGCACAGCCGTCCCCCCCTCCCTTCGCCAGCCTCGCTGAGGAGGAGGAGGAGGAGGTGGAAGAGCAGCGAGCGCGGAAG    186

ATG GAG GTC TCC ACG GAC CAA CCG CGG TGG GTG AGC CAC CAC CAC CCG GCC GTG CTC AAT GGG CAG CAC CCG GAC TCC CAC CAC    270
 M   E   V   S   T   D   Q   P   R   W   V   S   H   H   H   P   A   V   L   N   G   Q   H   P   D   S   H   H
CCC ACT CTT GGC CAT ACC TAC ATG GAC CCC ACG CAG TAT CCT CTC GCC GAG GAA GTG GAT GTA CTT TTT AAT ATC GAC GGA CAA    354
 P   T   L   G   H   T   Y   M   D   P   T   Q   Y   P   L   A   E   E   V   D   V   L   F   N   I   D   G   Q
GGC AAC CCC GTG CCT CCG TAT TAC GGC AAC TCC GTG CGA GCC ACC GTG CAG CGG TAC CCC ACG GCC CAT CAC GGC AGC CAG GTG    438
 G   N   P   V   P   P   Y   Y   G   N   S   V   R   A   T   V   Q   R   Y   P   T   A   H   H   G   S   Q   V
TGC CGG CCC CCT CTG CTG CAT GGT TGG CTG CCG TGG CTG GAC GGG AGC AAA GCG CTG AGC AGC CAC CAC AGC GCT TCC CCT TGG    522
 C   R   P   P   L   L   H   G   S   L   P   W   L   D   G   S   K   A   L   S   S   H   H   S   A   S   P   W
AAC CTC AGC CCT TTT TCC AAG ACC TCC ATC CAT CAC AGC TCG CCG GGA CCC CTC TCC GTT TAC CCG CCT GCT TCT TCC TCC ACT    606
 N   L   S   P   F   S   K   T   S   I   H   H   S   S   P   G   P   L   S   V   Y   P   P   A   S   S   S   T
TTA TCC GCC GGT CAC TCC AGC CCG CAC CTT TTC ACC TTC CCA CCG ACC CCT CCT AAA GAT GTG TCC CCG GAT CCG TCC ATC TCC    690
 L   S   A   G   H   S   S   P   H   L   F   T   F   P   P   T   P   P   K   D   V   S   P   D   P   S   I   S
ACT CCT GGC TCC ACC GGC TCC ACC CGG CAG GAT GAG AAG GAA TGC ATC AAA TAT CAG GTG TCC CTG GCT GAC ACC ATG AAG CTG    774
 T   P   G   S   T   G   S   T   R   Q   D   E   K   E   C   I   K   Y   Q   V   S   L   A   D   T   M   K   L
GAG TCG TCT CAC TCT AGG AGC AGC ATG GCC TCG TTA GGA GGA GCC ACC TCC TCC GCT CAT CAC CCC ATC ACT ACT TAC CCA CCG    858
 E   S   S   H   S   R   S   S   M   A   S   L   G   G   A   T   S   S   A   H   H   P   I   T   T   Y   P   P
TAT GTC CCC GAA TAT AGC TCT GGA CTT TTT CCC CCC AGC AGC CTC TTA GGA GGA TCG CCT ACC GGG TTC GGG TGC AAA TCA CGA    942
 Y   V   P   E   Y   S   S   G   L   F   P   P   S   S   L   L   G   G   S   P   T   G   F   G   C   K   S   R
CCG AAA GCG CGG TCG AGC ACA GAA GGC AGG GAG TGT GTA AAT TGT GGG GCT ACC TCA ACC CCT CTG TGG AGA AGA GAC GGC ACC   1026
 P   K   A   R   S   S   T   E   G   R   E   C   V   N   C   G   A   T   S   T   P   L   W   R   R   D   G   T
GGG CAC TAC TTG TGT AAC GCC TGT GGA CTC TAT CAC AAA ATG AAT GGG CAG AAC CGA CCC CTG ATT AAA CCC AAG AGA AGG CTG   1110
 G   H   Y   L   C   N   A   C   G   L   Y   H   K   M   N   G   Q   N   R   P   L   I   K   P   K   R   R   L
TCT GCA GCC AGG AGG GCG GGC ACG TCC TGT GCT AAC TGT CAG ACC ACC ACC ACC ACC CTG TGG AGG AGA AAT GCC AAC GGC GAT   1194
 S   A   A   R   R   A   G   T   S   C   A   N   C   Q   T   T   T   T   T   L   W   R   R   N   A   N   G   D
CCT GTC TGT AAT GCC TGT GGG CTC TAC TAC AAG CTG CAC AAT ATT AAC AGA CCC CTG ACT ATG AAG AAA GAA GGA ATT CAG ACC   1278
 P   V   C   N   A   C   G   L   Y   Y   K   L   H   N   I   N   R   P   L   T   M   K   K   E   G   I   Q   T
AGA AAC CGA AAA ATG TCT AGC AAA TCC AAA AAG TGC AAA AAG GTC CAT GAC AAC CTT GAA GAC TTT CCC AAG AGC AGC TCC TTT   1362
 R   N   R   K   M   S   S   K   S   K   K   C   K   K   V   H   D   N   L   E   D   F   P   K   S   S   S   F
AAC CCC GCC GCC CTC TCC AGA CAC ATG TCC TCT ATC AGC CAC ATT TCA CCC TTC AGT CAC TCC AGC CAC ATG CTG ACT ACA CCG   1446
 N   P   A   A   L   S   R   H   M   S   S   I   S   H   I   S   P   F   S   H   S   S   H   M   L   T   T   P
ACA CCG ATG CAT CCT CCA TCC AGC CTC TCC TTT GGA CCT CAC CAT CCC TCC AGC ATG GTC ACT GCC ATG GGT TAG TGAGAGACTCC   1532
 T   P   M   H   P   P   S   S   L   S   F   G   P   H   H   P   S   S   M   V   T   A   M   G   *

CCTGCTGAATGCTTACGGTCTCAAAATGAGATTTACTTTATATACTTGCATTTTTGCAGGCGTGTGGTTATGGGTCTGATGTCCCAAATCAAATGGGAGGGGGAAAAAGAA   1643
AAAAAAAAGAATAAAAAGAATAAAAATAAAAATGTTATTTGAAGGCGTAAGAGCAAAAAAAATAATAATAAAAAAAGAAAAAAACTACAAAAAGCACAAAAAAGGAAAAAA   1754
AAAGAAAAAAAAAAAAAAAA                                                                                                 1774
```

Figure 3. Sequence of NF-E1c cDNA. The nucleotide sequence of cDNA clone p31a is shown; numbers at *right* represent nucleotides from the 5′ end of the insert. Predicted protein sequence of the longest ORF is shown below the nucleic acid sequence. The shaded area corresponds to the finger domain and sequences encoding it.

legend to Fig. 5). Thus, although it is not a prevalent species, NF-E1b is clearly not a rare mRNA in these cells.

The tissue distribution of expression of the third gene, NF-E1c, is once again distinct when compared to that of the other NF-E1 mRNAs. Like NF-E1b, NF-E1c is present in definitive reticulocytes at much lower relative concentrations than the extremely abundant chicken NF-E1a mRNA (Fig. 5C). Similarly, the 3.3-kb mRNA encoding NF-E1c is found in adult kidney and abundantly in embryonic brain. NF-E1c is unique among the gene family in that it is the only member whose expression is restricted to a distinct subset of erythroid cells, found only in definitive (adult) reticulocytes. It is also the sole member of the family expressed in MSB-1 (T-lymphocyte) cells (Akiyama and Kato 1974; Beug et al. 1981).

To examine the expression of these factors specifically within hematopoietic lineages and during maturation of hematopoietic progenitors, we performed further RNA blot analysis of the NF-E1 cDNAs to mRNAs derived from cells transformed with avian viruses. In this series of experiments, RNA from equal numbers of viable cells was electrophoresed in each lane (Fig. 5D–F), rather than equivalent amounts of RNA (Fig. 5A–C). Note that because the total RNA recovered per cell decreases precipitously during erythrocyte maturation (Fig. 5D–F, lanes 12–16; H. Beug and J.D. Engel, data not shown),

```
GF-1     MDFPGLGALG -TSEPLPQFV DSALVSSPSD STGFFSSG-- --PEGLDAAS SSTSP-N-AA   53
NF-E1a   MEFVALGGPD AGS-PTP-F- ---------PD EAGAFL---- ----GLGGGE RTEA------   35
NF-E1b   ME-VATDQPR WMTHH---AV ---LNGQHPE SHHPGLAHNY MEPAQLLPPD EVDVFFNHLD   53
NF-E1c   ME-VSTDQPR WVSHHHP-AV ---LNGQHPD SHHPTLGHTY MDPTQYPLAE EVDVLFN-ID   54

GF-1     TAAASALAYY RE-AEAYRHS PVFQVYP--- ---------- --LLNSMEGI P---------   88
NF-E1a   ---------- ---------G GLLASYP--- ---------- --PSGRVSLV P---------   52
NF-E1b   SQGNP---YY ANSAHA-RAR -VS--YSQAH ARLTGSQMCR PHLIHS-PGI PWLDSSKAAL  105
NF-E1c   GQGNPVPPYY GNSV---RAT -VQR-YPTAH HGSQV---CR PPLLHG--SL PWLDGSKA-L  103

GF-1     ---------G GSPY------ --ASWAYGKT ALYPA---ST VCPSHEDAPS QA--------  120
NF-E1a   ---------- ---------- ----WADTGT LGTPQ----W VPPATQMEPP H---------   75
NF-E1b   SAHHHN-PWT VNPFTKTPLH PSAAGAPGAI SVYPG--SST SSTASVSSLT PASHSGSHLF  162
NF-E1c   SSHHSASPWN LSPFSKTSIH HSSPGP---L SVYPPASSST -LSA------ --GHSSPHLF  151

GF-1     ---------- ---------- ---------L EDQEG-KSNN TFL-DTLKTE RLSPDL---L  146
NF-E1a   ---------- ---------- ---------- ---------- -YL-ELLQPP RGSPPHPSS-   92
NF-E1b   GFPPTPPKEV SPDPNSTSAA SPSSSAGARQ EDKDSIKYQV S-LSEGMKME SASP-LRSSL  220
NF-E1c   TFPPTPPKDV SPDPSISTPG STG---STRQ DEKECIKYQV S-LADTMKLE S-SH-SRSSM  205

GF-1     T-LGTALPAS ----LPVTGS AYGGA--DFP SPFFSPTGSP LSSAAYSS-- PKFHGSLPLA  197
NF-E1a   ---GPLL--- ---------- ---------- ---------P LSS------- ------GPP-  103
NF-E1b   TSMG-AQP-S THHPIP-TYP SYVPAAHDYS SSLFHP-GSF LGGPA-SSFT PKPR-SKARS  274
NF-E1c   ASLGGATS-S AHHPIT-TYP PYVP---EYS SGLFPP-SSL LGGSP-TGFG CKSR-PKARS  257

GF-1     PCEARECVNC GATATPLWRR DRTGHYLCNA CGLYHKMNGQ NRPLIRPKKR MIVSKRAGTQ  257
NF-E1a   PCEARECVNC GATATPLWRR DGTGHYLCNA CGLYHRLNGQ NRPLIRPKKR LLVSKRAGTV  163
NF-E1b   CSEGRECVNC GATATPLWRR DGTGHYLCNA CGLYHKMNGQ NRPLIKPKRR LSAARRAGTC  334
NF-E1c   STEGRECVNC GATSTPLWRR DGTGHYLCNA CGLYHKMNGQ NRPLIKPKRR LSAARRAGTS  317

GF-1     CTNCQTTTTT LWRRNASGDP VCNACGLYFK LHQVNRPLTM RKDGIQTRNR KASGKGKK-K  316
NF-E1a   CSNCQTSTTT LWRRSPMGDP VCNACGLYYK LHQVNRPLTM RKDGIQTRNR KVSSKGKK-R  222
NF-E1b   CANCQTTTTT LWRRNANGDP VCNACGLYYK LHNVNRPLTM KKEGIQTRNR KMSNKSKKSK  394
NF-E1c   CANCQTTTTT LWRRNANGDP VCNACGLYYK LHNINRPLTM KKEGIQTRNR KMSSKSKKCK  377

GF-1     RGSNLAGAGA AEGPAGGFMV VAGSSSSGNC GEVASGLALG TAGTAHLYQG LGPVVLSGPV  376
NF-E1a   RPP---GGGN PSATAGGGAP MGGGGDPSM- PPPPPPPAAA PPQSDALY-A LGPVVLSG--  275
NF-E1b   KGS-----EC FEELSK-CMQ EK--SSPFS- -------AAA LA-S-HMA-P MG--------  427
NF-E1c   KVH-----DN LEDFPK---- ----SSSFN- -------PAA L--SRHMS-S I---------  404

GF-1     SHLMPFPGP- -LLGSPTTSF PTGPA-PTTS STSVIAP-LS S                    413
NF-E1a   -HFLPF---- ---GNSGGFF GGGAG-GYTA P-----PGLS PQI                  304
NF-E1b   -HLPPFSHSG HIL--PT--- PT-PIHP-SS SISFGHPHPS SMVTAMG              466
NF-E1c   SHISPFSHSS HML---TT-- PT-PMHPP-S SLSFGPHHPS SMVTAMG              444
```

Figure 4. Sequence comparison of the chicken NF-E1 and murine GF-1 proteins. Sequence alignment for maximum identity between the murine GF-1 and the chicken NF-E1a, NF-E1b, and NF-E1c proteins was performed by pairwise matching of all six possible combinations with the resident sequence analysis software in the IBI gel analysis system and then by visual matching. Only identities to the GF-1 protein (Tsai et al. 1989) are shaded. Dashes indicate breaks used to achieve maximum sequence identity for all four proteins. Numbers at *right* correspond to the last amino acid shown on each line.

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

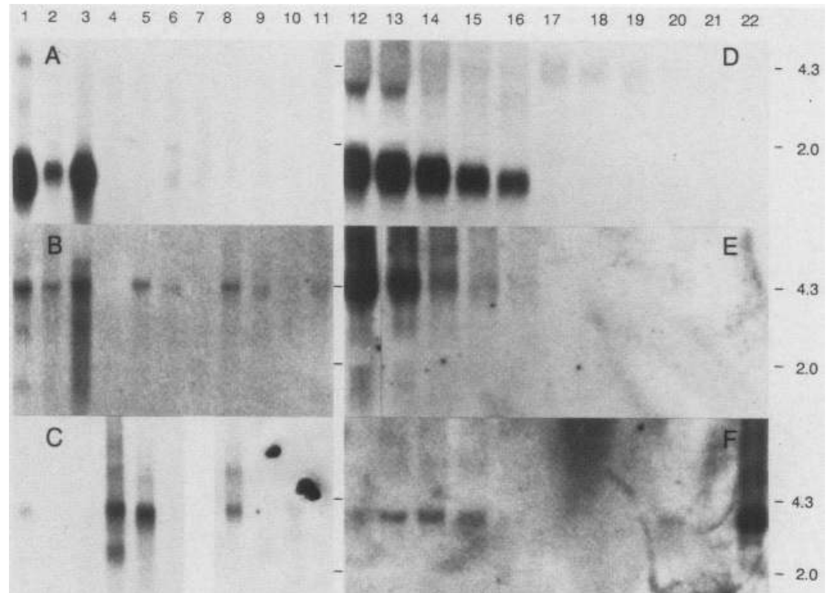


Figure 5. Expression of NF-E1 mRNAs. RNAs isolated from various chicken cell lines, virally transformed cell clones, or specific tissues were denatured, electrophoresed, and blotted to nylon membranes, as described previously (Riddle et al. 1989). The probes used were full-length NF-E1a (*A* and *D*), NF-E1b (*B* and *E*), or NF-E1c (*C* and *F*) cDNAs (Fig. 1). All blots were washed at moderate stringency (120 mM monovalent cation, 55°C). (*A–C*) Two micrograms of poly(A)$^+$ RNA electrophoresed in each lane was isolated from anemic adult hen (definitive) reticulocytes (lanes *1*); HD6 erythroid progenitor cells (lanes *2*); embryonic (primitive) reticulocytes from 4.5-day embryos (lanes *3*); Marek's virus-transformed chicken T-lymphoma cells (lanes *4*; MSB-1; Akiyama and Kato 1974); 11-day chick embryo brain (lanes *5*); 11-day chick embryo liver (lanes *6*); adult chicken perfused liver (lanes *7* in *A* and *B*, not shown in *C*); adult chicken perfused kidney (lanes *8*); 11-day chick embryo cardiac muscle (lanes *9*); 11-day chick embryo skeletal muscle (lanes *10*); and 11-day chick embryo fibroblasts (lanes *11*). (*D–F*) Total cellular RNA (representing 6.5×10^6 viable cell equivalents of RNA in each lane; see Results) was isolated from chicken erythroblasts transformed by a recombinant virus expressing the ts21 v-*myb* and human epidermal growth factor receptor (hEGFR) genes (Khazaie et al. 1988; H. Beug, unpubl.), either containing hEGF and cultured at 37°C (lanes *12*) or induced to differentiate by depletion of hEGF and grown for 1 day (lanes *13*), 2 days (lanes *14*), 3 days (lanes *15*), or 4 days (lanes *16*) at 42°C in the presence of anemic chicken serum and concentrated supernatant (REV) factor derived from NPB4 cells (Zenke et al. 1988); ts21 E26-transformed myeloid cells (Beug et al. 1984) shifted to the nonpermissive temperature for 0 days (lanes *17*), 2 days (lanes *18*), or 4 days (lanes *19*); a clone of v-*rel*-transformed pre-B/pre-T lymphoblasts (P. Hayman, unpubl.; lanes *20*); RP-9 cells (a transformed B-lymphocyte cell line; Beug et al. 1981; H. Beug, unpubl.; lanes *21*); and MSB-1 (T-lymphoma) cells (Beug et al. 1981; H. Beug, unpubl.; lanes *22*). After hybridization to individual random primer-labeled NF-E1 cDNA clones, each of the blots was stripped and rehybridized to a mixture of cDNA probes encoding the chicken β-actin and erythrocyte band 3 proteins (Cleveland et al. 1980; Kim et al. 1989). The resultant autoradiographs (not shown) were then used to determine the integrity, size, and relative concentration of the samples, whereas the band 3 hybridization signal was additionally used to estimate the degree of RBC contamination in the various nonerythroid chicken tissues (lanes *4–11*). All of the tissue RNAs (lanes *1–11*) contained equivalent β-actin mRNA signals (to within a factor of two), and none of the mRNAs were significantly contaminated (>5%) with RBCs. Size markers shown at *right* correspond to the positions of β-actin (2.0 kb) and erythrocyte band 3 (4.3 kb) mRNAs. Exposure times were (*A*) 40 min; (*B* and *E*) 48 hr; (*C* and *F*) 18 hr; (*D*) 3 hr.

one cannot directly compare the results of Figure 5, A–C, with those shown in Figure 5, D–F.

NF-E1a is not found in lymphoid lineages or in tsE26-transformed myeloid cells; it is abundantly expressed only within the erythroid lineage (Fig. 5D). NF-E1b is expressed at relatively high levels in immature erythroid cells but decreases sharply in abundance (on a per cell basis) during definitive erythroid cellular maturation and is not expressed in myeloid or lymphoid cell lineages (Fig. 5E). Unlike NF-E1a, NF-E1c is not expressed in immature hEGFR/ts *myb*-transformed erythroblast-like cells and is strongly up-regulated during definitive erythroid cellular maturation (Fig. 5F). In addition, NF-E1c is expressed in v-*rel*-transformed lymphoid cells (a pre-B/pre-T precursor cell; P. Hayman and H. Beug, unpubl.) at approximately the same abundance as in partially differentiated erythroblasts (Fig. 5F, lanes 13 and 20) but accumulates to ~20-fold greater abundance in MSB-1 cells. The abundant expression of this factor in T lymphocytes may be of functional significance (see Discussion).

Sequence-specific DNA binding properties of NF-E1 proteins

NF-E1b and NF-E1c are predicted to encode finger domains with >90% identity to the DNA binding domains of the murine factor GF-1 and chicken NF-E1a (Evans and Felsenfeld 1989; Tsai et al. 1989), suggesting that the NF-E1b and NF-E1c proteins would also bind the WGATAR motif with high affinity. To compare the DNA binding properties of the NF-E1 proteins, each was expressed by in vitro transcription and translation. Single, full-length transcripts were produced (Fig. 6A) and used to direct protein synthesis of the NF-E1 factors in rabbit reticulocyte lysate (see Methods). Proteins of ~39.5, 56, 55, and 51 kD were obtained for NF-E1a, NF-E1b, NF-E1c, and GF-1, respectively (Fig. 6B). These are 6–8 kD larger than expected on the basis of conceptual translation of the cDNA clones (Fig. 4), possibly as a result of the high proline content of the proteins (Hope and Struhl 1985).

NF-E1a, NF-E1b, NF-E1c, and GF-1 proteins synthe-

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

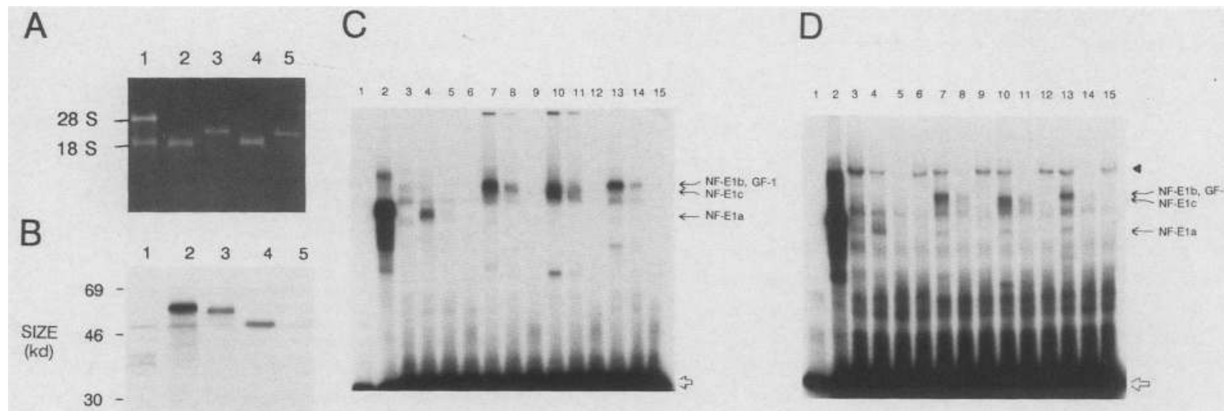


Figure 6. NF-E1 proteins expressed in vitro specifically bind to the WGATAR consensus sequence. (*A*) In vitro transcription of NF-E1a, NF-E1b, NF-E1c, and GF-1 cDNA clones. RNA products of bacteriophage polymerase-initiated transcription were electrophoresed on a 1.5% agarose gel containing formaldehyde. (Lane *1*) 0.46 μg of total anemic hen RBC RNA. The positions of the 18S and 28S rRNA bands are indicated. (Lanes *2–5*) Two and one-half microliters of each of the in vitro transcription reactions with the following cDNA templates: (lane *2*), NF-E1a; (lane *3*) NF-E1b; (lane *4*) NF-E1c; (lane *5*) GF-1. (*B*) In vitro translation of NF-E1a, NF-E1b, NF-E1c, and GF-1 RNAs. RNA templates were translated in rabbit reticulocyte lysate in the presence of [^{35}S]methionine. Two microliters of translation reactions was run on a 15% SDS–polyacrylamide gel and subjected to fluorography; molecular size standards are as indicated at *left*. (Lane *1*) NF-E1a; (lane *2*) NF-E1b; (lane *3*) NF-E1c; (lane *4*) GF-1; (lane *5*) no added RNA. (*C* and *D*) Gel mobility-shift assays using NF-E1a, NF-E1b, NF-E1c, and GF-1 in vitro translation products. The in vitro translation reaction (1.5 μl) was used to assay the binding properties of each NF-E1 protein with the MαP (*C*) or CβE (*D*) ^{32}P-labeled probes (see Results). Binding reactions were run on 5% nondenaturing polyacrylamide gels. Nonspecific poly[d(I-C)]/[d(A-T)] competitor was used at 2 μg for basic extract (see Methods) or 50 ng for in vitro translation products. Unlabeled CβE (25 ng) or MαP (50 ng) oligonucleotides were added as indicated: (lane *1*) free probe; (lane *2*) 0.5 μl of definitive erythrocyte basic extract (Gallarda et al. 1989); (lane *3*) no RNA added to the in vitro translation reaction; (lanes *4–6*) NF-E1a; (lanes *7–9*) NF-E1b; (lanes *10–12*) NF-E1c; (lanes *13–15*) GF-1. In vitro translation products were incubated with radiolabeled NF-E1 oligonucleotide plus added nonspecific competitor (lanes *4*, *7*, *10*, and *13*), a 50-fold excess of unlabeled CβE oligonucleotide (lanes *5*, *8*, *11*, and *14*), or a 100-fold excess of unlabeled MαP oligonucleotide (lanes *6*, *9*, *12*, and *15*). The arrows indicate high-affinity complexes of NF-E1 proteins with the probes. The open arrow shows the position of migration of free probe (*D*) the arrowhead indicates a complex formed by a protein present in rabbit reticulocyte lysate that preferentially binds to the (dimer site) CβE probe (see Results).

sized in vitro were tested in gel mobility-shift assays with two different probes. One, from the mouse α-globin promoter (MαP; Tsai et al. 1989), contains a single NF-E1 binding site (TGATAA), whereas the other, from the chicken β-globin enhancer (CβE; Gallarda et al. 1989), contains two binding sites, AGATAA and TGATAG, in inverted orientation, as encountered in the natural β-globin enhancer (Choi and Engel 1986; Hesse et al. 1986). All of the chicken NF-E1 or murine GF-1 proteins formed unique complexes with both probes in the presence of 100-fold excess of nonspecific competitor DNA. These complexes could be competed with either a 50-fold (CβE) or 100-fold (MαP) excess of unlabeled specific competitor (Fig. 6C,D).

One also detects several complexes formed between the probes and rabbit NF-E1 proteins in the reticulocyte lysate; these can be categorized into two types. One type of complex was found with either probe and could be competed by either of the unlabeled NF-E1 sequence oligonucleotides; there appear to be three such complexes formed (e.g., Fig. 6C, lanes 3 and 6). In contrast, a second, lower mobility complex was formed only with the CβE probe (Fig. 6D). This complex, which is competed by CβE but not by MαP, possibly represents cooperative dimer binding that could occur on the former, but not on the latter (single site) oligonucleotide.

Gel mobility shift assays were also performed with amino-terminally truncated NF-E1b and NF-E1c proteins. Deletion of the first 260 amino acids of the NF-E1b protein or the first 131 amino acids of the NF-E1c protein still permits formation of high-affinity complexes with the CβE probe. In the case of NF-E1b, deletion of the amino-terminal region to within 20 amino acids of the finger domain still allows specific complex formation. With further truncation (deleting the entire finger domain), binding was completely eliminated (data not shown). Thus, specific binding of the chicken NF-E1 proteins to the WGATAR consensus sequence requires the finger domain.

Trans-activation by NF-E1 proteins

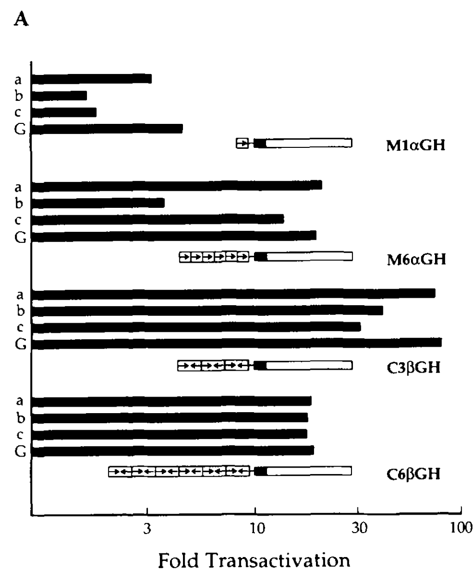
The ability of the NF-E1 proteins to stimulate transcription in vivo was examined by using a cotransfection *trans*-activation assay (Giguere et al. 1986). In these experiments, the transcriptional activity of a reporter gene [human growth hormone (hGH)] was directed by the rabbit β-globin TATA box and either one or six copies of the MαP oligonucleotide or three or six copies of the CβE oligonucleotide (either 6 or 12 copies of the binding site); these are subsequently referred to as M1αGH, M6αGH, C3βGH, and C6βGH, respectively. In each

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

assay, a plasmid containing one of the NF-E1 cDNA clones [transcriptionally directed by the Rous sarcoma virus long terminal repeat (RSV LTR) in plasmid TFAneo (Federspiel et al. 1989)] was included as a potential *trans*-activator. Activation of the hGH constructs was assayed by measuring the amount of hGH peptide secreted into the tissue-culture medium by radioimmunoassay (RIA; see Methods). Recipient cells used in these experiments were the human HeLa cell line, NIH-3T3 mouse fibroblasts, and immortal quail (QT6) fibroblasts (Scherer et al. 1953; Todaro and Green 1963; Moscovici et al. 1977).

The *trans*-activation activities are quite different, depending on the recipient cell origin, as well as the number and position of NF-E1 binding sites incorporated into the reporter gene plasmids. The murine GF-1 protein stably accumulates in all transfected cells and is able to *trans*-activate all hGH reporter gene constructs. In contrast, the chicken NF-E1 proteins fail to significantly accumulate in 3T3 or HeLa cells (D.I.K. Martin and S.H. Orkin, unpubl.) but are stable in QT6 (Fig. 7B). Because of species-dependent differences in the stability of the various *trans*-activating proteins, we have drawn conclusions based only on the results derived from the QT6 cell transfections.

All of the hGH reporter gene plasmids are *trans*-activated by the murine (GF-1) factor and each chicken NF-E1 protein in QT6 cells (Figure 7A); however, the levels of activity in this assay quantitatively differ depending on whether the reporter contains 1, 6, or 12 binding sites for the factor. Thus, *trans*-activation of M6αGH by NF-E1a is ~5-fold greater than the activation of M1αGH, whereas the level of activation of M6αGH by NF-E1c is ~10-fold greater than the same protein acting on the M1αGH reporter. Clearly, equal numbers of binding sites in the plasmids are not the sole determinant for efficient *trans*-activation: M6αGH and C3βGH each contain six binding sites, but the resultant *trans*-activation with each NF-E1 protein differs markedly. Curiously, activation of C6βGH was identical (within experimental error) for all four proteins tested and lower than achieved with C3βGH. Thus, one may conclude that all of the chicken NF-E1 DNA binding proteins are also *trans*-acting factors and that they are capable of distinguishing between various templates in vivo.



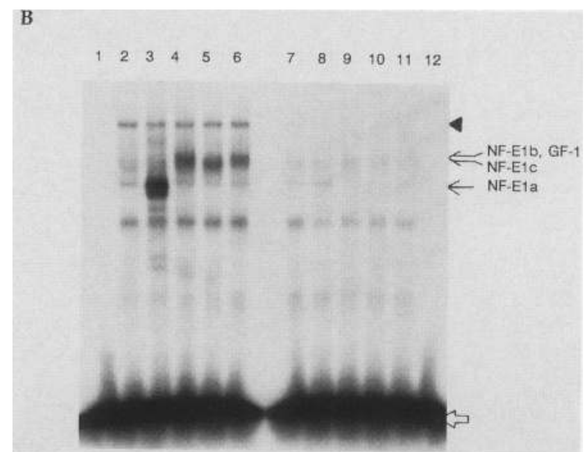


Figure 7. Transcriptional activation by NF-E1 proteins. (*A*) The activation assays were performed by transfecting reporter plasmids M1αGH, M6αGH, C3βGH, or C6βGH (8 μg each) into QT6 cells with TFAneo plasmids (2 μg each), expressing either the NF-E1a, NF-E1b, NF-E1c, or murine GF-1 proteins (see Methods). Levels of *trans*-activation were calculated as described in Methods. After supernatants were collected for RIA assay of secreted hGH, the transfected cells were scraped from the plates and cell lysates were prepared (Tsai et al. 1989). Five microliters of the cell lysate (of 25 μl/10^6 cells) was used in gel shift analysis as described in the legend to Fig. 6, except the binding reactions were carried out on ice for 15 min. Nonspecific poly[d(I-C)]/[d(A-T)] competitor was used at 2.25 μg for all reactions, and unlabeled MαP competitor (50 ng) was added to the reactions in lanes *7–12*. Cell lysates were from TFAneo (vector) transfection (lanes *2* and *7*); TFAneo/NF-E1a transfection (lanes *3* and *8*); TFAneo/NF-E1b transfection (lanes *4* and *9*); TFAneo/NF-E1c transfection (lanes *5* and *10*); TFAneo/GF-1 transfection (lanes *6* and *11*). The position of migration of free probe (lanes *1* and *12*; open arrow), specific complexes (arrows), and QT6-specific NF-E1 specific complexes (arrowhead) are indicated.

Discussion

NF-E1 is a developmentally regulated multigene family

We report the physical and biological characterization of a group of related *trans*-acting factor proteins whose mRNAs are expressed in a variety of chicken cell types. The proteins examined in this study all bind to a common consensus DNA motif WGATAR, bind with high affinity to this motif in vitro (Fig. 6), and can serve as potent positive *trans*-acting factors in vivo (Fig. 7).

When compared to the size of the cloned cDNA segments (Figs. 2 and 3) and to the size of the intact mRNAs derived from the RNA blots (Fig. 5) and primer extension data (not shown), it appears that the entire 3′ end of NF-E1a is represented; whereas both NF-E1b and

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

NF-E1c are lacking sequence representation from both the 5'- and 3'-untranslated regions. Although NF-E1a lies in a relatively poor context for translation, the longest clone is missing only 39 nucleotides from the 5' end (M.W. Leonard, unpubl.). NF-E1b is in a good consensus sequence translation context and matches the predicted amino terminus of the NF-E1c protein with a high degree of amino acid sequence identity (Fig. 4). NF-E1c is in a reading frame preceded by an in-frame stop codon 183 nucleotides 5' to the assigned initiation codon. In summary, from these and the *trans*-activation data, we conclude that the full coding sequence for each of the NF-E1 proteins is represented in Figure 4.

The physical properties of the factor NF-E1a (Eryf1) have been well characterized previously (Evens and Felsenfeld 1989). Here, we show that this DNA binding protein is also an effective *trans*-acting factor in vivo (Fig. 7). NF-E1a mRNA (and protein; M. Yamamoto, unpubl.) is abundantly and exclusively expressed in erythroid cells (Fig. 5) and therefore might, as its murine counterpart GF-1, serve as the primary determinant for genes that are to be expressed in the erythroid lineage. However, because it is equally abundant in primitive and definitive erythroid cells, it is unlikely that NF-E1a is also a primary determinant of hemoglobin switching, as chicken erythroid cells express different β-globin isotypes during development (Bruns and Ingram 1973).

Factor NF-E1b appears to be the most enigmatic of the three factors reported here. NF-E1b is clearly restricted to expression in distinct, developmentally unrelated cell types (Fig. 5); however, it is expressed at relatively high levels in immature definitive erythroid cells and declines very rapidly in abundance as these cells continue to mature (cf. the relative decrease in NF-E1a mRNA levels in Fig. 5D to NF-E1b mRNA levels in Fig. 5E, lanes 12–16). Because NF-E1a increases in relative poly(A)$^+$ mRNA abundance during the same time frame (Fig. 5A, lanes 1 and 2), it seems possible that NF-E1b may be the very earliest form of a *trans*-acting factor capable of activating the WGATAR consensus motif during early erythropoiesis, and is subsequently replaced by a "late" factor, NF-E1a or NF-E1c. NF-E1b is also the only factor in which we have been able to detect tissue-specific RNA splicing (M.W. Leonard and M. Yamamoto, unpubl.).

NF-E1c is abundant in embryonic brain and mature T-lymphoid cells but is not expressed in embryonic (primitive) or immature definitive erythroid cells (Fig. 5C,F). Although the role of this factor in erythropoiesis is unclear, it appears to be the only member of this family expressed almost exclusively in a different hematopoietic lineage. An indication as to the part this factor may play in T-cell-specific transcriptional regulation is suggested by recent observations that show detailed tissue-specific footprints in the human T-cell receptor δ gene enhancer, which lie directly over a consensus NF-E1-binding motif (the TGATAA repeat comprising footprint δE4; Redondo et al. 1990). Furthermore, this NF-E1 consensus has been identified in other gene regulatory regions important for T-cell-specific expression, for example, within the murine and human T-cell receptor α gene enhancers [AGATAG comprising footprint NFα1 (Winoto and Baltimore 1989) and AGATAG within footprint Tα3 (Ho et al. 1989)] and at four locations within the human T-cell leukemia virus-III (HTLV-III) LTR U3 region (Rosen et al. 1985).

Why do erythroid cells have more than one NF-E1 protein?

The mRNAs for the three chicken proteins characterized here are all expressed in definitive erythroid cells, albeit at vastly different abundances; thus, one might infer that each plays some role in transcriptional regulation in erythroid cells. One puzzling aspect of these studies is why there are three different *trans*-acting factors that recognize the same consensus motif, with similar affinities, within a single cell. Several alternatives are apparent.

One possibility is that the various NF-E1 proteins differ in their *trans*-activation potential, as has been reported recently for the ubiquitous Oct-1 and lymphoid-specific Oct-2 factors (Tanaka and Herr 1990), and is suggested by the amino acid sequence divergence among these factors outside the DNA binding domain. A second possibility for why factors with a presumably identical DNA binding specificity may exist in the same cells is that the DNA sequence of any consensus binding site may be very important in determining which factor within a multigene family is actually bound to that site. One might predict that the ambiguities in accumulating a consensus sequence (e.g., whether a T or A residue is encountered at the first position of the NF-E1 consensus WGATAR) are not a reflection of a degenerate code to which several factors can bind with equal affinity but, rather, dictate a strong preference for which factor in a multigene family is actually used at a given site. Recent observations in the analysis of GF-1 (Plumb et al. 1989; D.I.K. Martin and S.H. Orkin, in press) and a more distantly related homolog (Wilson et al. 1990) support this hypothesis.

Another possibility addresses the more perplexing and currently only poorly understood nature of what arrangement of protein/DNA binding activities constitute a genuine enhancer. Clearly, multiple consensus sequence binding sites for these factors are not sufficient to form a functional enhancer: Whereas multiple copies of the NF-E1 consensus sequence inserted directly 5' to the human β-globin TATA box are sufficient to stimulate high levels of transcription in the presence of any of these factors (Fig. 7A), if the binding sites are placed 3' to a reporter gene, transcriptional stimulation is dramatically reduced (S.H. Orkin and J.D. Engel, data not shown). Thus, an arrangement of simple multimeric binding sites for the NF-E1 proteins bear a much stronger resemblance to upstream promoter elements than to genuine enhancers, because the latter can be found at far distance locations relative to the promoters they activate (Maniatis et al. 1987).

Because enhancers are complex modules, consisting of

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

tightly clustered tissue-specific and constitutive *trans*-acting factor binding sites, it may be that none of the artificial constructs examined in cotransfection *trans*-activation assays are appropriate models for reconstitution or reflection of the activity of a genuine enhancer. One can imagine that not only is the precise DNA sequence of a factor binding site important, but the sequence context within these modules (i.e., the nearest adjacent proteins themselves) could be of equal or greater importance. Thus, the precise stereochemical relationships determined by sequence context could dictate binding of adjacent factors in a crowded complex. In this model, perhaps the overall sequence context is the major determinant of which factor might bind a given motif within a complex module when several members of a multigene family are expressed in one cell type.

How might this hypothesis aid in explaining why multiple NF-E1 factors exist in chicken erythroid cells? Consistent with the observed mRNA expression profile (Fig. 5), the presence of NF-E1b in primitive erythrocytes could indicate that this is the family member bound within the restricted confines of the shared β/ϵ-globin enhancer in vivo (Choi and Engel 1988; Nickol and Felsenfeld 1988) and is involved in directing ϵ-globin transcription in primitive cells. During late definitive erythropoiesis, NF-E1c may then replace NF-E1b binding to the enhancer and, in conjunction with additional *trans*-acting factors such as NF-E4 (Gallarda et al. 1989), direct adult β-globin transcription. A similar model has recently been proposed to account for the observation that the factors C/EBP and DBP, which bind to the same DNA motif within the rat albumin promoter, are both expressed in hepatocytes, albeit with distinct developmental profiles (Mueller et al. 1990). Because the two minor abundance NF-E1 factors are so structurally dissimilar to NF-E1a outside of the DNA binding domain and yet so similar to one another (Fig. 4), perhaps the nearest neighbor protein–protein interactions within the tightly clustered confines of the β/ϵ-globin enhancer allow only NF-E1b and NF-E1c to bind to this particular site in vivo. The cloning of three members of the NF-E1 multigene family should allow direct assessment of the several possibilities raised here.

Methods

cDNA cloning and DNA sequencing

Two independent λgt11 cDNA libraries were used to isolate chicken NF-E1 cDNA clones. The BV4 library (Sap et al. 1986) was prepared with polyadenylated RNA isolated from a pool of 10-day-old chicken embryos. The RBC4 library was prepared in a new λgt11 derivative vector λgt11d (for directional; marketed as λgt11 Sfi-Not, Promega, Madison, WI), developed to promote oriented insertion of cDNA libraries. The vector was made by digestion of λgt11 with *Eco*RI and ligation to a linker that destroyed the original *Eco*RI site and simultaneously created new sites within the vector (5′ to 3′ in *lacZ* sense) for *Sfi*I, *Eco*RI, and *Not*I:

5′-AATTTCACGGCCATCATGGCCGTGAATTCCATAGCGGCCGCCTCTA
AGTGCCGGTAGTACCGGCACTTAAGGTATCGCCGGCGGAGATTTAA-5′
*Sfi*I *Eco*RI *Not*I

First-strand cDNA synthesis was initiated by using an oligonucleotide primer with a *Not*I site at the 5′ end, followed by 12 deoxythymidylate residues. Upon completion of double-stranded cDNA synthesis by standard procedures (Gübler and Hoffman 1983), *Eco*RI adapters were added. The cDNA was digested with *Not*I and the electroeluted cDNAs (>1.2 kb in size) were then ligated to λgt11d, which had been digested with *Not*I and *Eco*RI. The initial complexity of the RBC4 library was 10^6 PFU.

By plaque hybridization (Benton and Davis 1977) with a polymerase chain reaction (PCR)-amplified DNA fragment corresponding to the DNA binding domain of the murine GF-1 sequence (nucleotides 481–980 of GF-1 cDNA; Tsai et al. 1989), 6×10^5 PFU from the RBC4 library were first screened. The screening was performed at room temperature in the presence of 50% formamide, and filters were washed at 37°C in a solution containing 1× blot wash [50 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 0.1% SDS] plus 1× SSC. After plaque purification, phage DNA was prepared by the method of Chisholm (1989). The BV4 library was then screened by using a cDNA fragment specifying the DNA binding domain of NF-E1b as a probe (corresponding to nucleotides 938–1542 of the NF-E1b sequence; Fig. 2). The low-stringency hybridization and washing conditions used in the previous screen were again employed. Finally, the BV4 library was rescreened by using both of the 5′-end fragments of cDNA clones λ10a and λ8a, the longest cDNAs representing loci NF-E1b and NF-E1c, respectively, from the first screen. The probes were hybridized at 42°C in the presence of 50% formamide, and the filters were washed at 42°C in 1× blot wash.

All independent cDNA clones were subcloned into either pGEM3 or pGEM7Zf(+) (Promega), and most of the independent subclones were sequenced directly from both ends. Selected subclones were subject to ordered serial deletion mutagenesis according to the method of Henikoff (1984). Nucleotide sequence of NF-E1a was determined from subclones p33d and p36c, that of NF-E1b from p10a and p30a, and NF-E1c from p8a and p31a. DNA sequencing was performed by using the dideoxy chain termination method (Sanger et al. 1977) with T7 DNA polymerase and alkali-denatured double-stranded plasmid DNA as template (Choi and Engel 1986). 7-Deaza-dGTP was used in place of dGTP (Mizusawa et al. 1986).

In vitro expression of NF-E1 cDNA clones

NF-E1a, NF-E1b, NF-E1c, and GF-1 cDNAs were subcloned into plasmid vectors for in vitro expression. A 1.0-kb *Not*I–*Eco*RI fragment of NF-E1a was inserted into the *Nco*I–*Eco*RI sites of vector pBSATG containing a strong translation initiation consensus sequence (Clerc et al. 1988). GF-1 cDNA was digested with *Xho*I, filled in by use of the Klenow fragment of DNA polymerase I to create blunt ends, and redigested with *Nco*I. The resultant 1.5-kb fragment of GF-1 was inserted into *Nco*I–*Sma*I sites of pBSATG. Both of these were prepared such that an ATG within the *Nco*I site (CCATGG) would initiate translation of the cDNA in the proper reading frame. A 2.4-kb *Nco*I–*Eco*RI fragment of NF-E1b was inserted into pGEM4 (Promega), as was a 1.8-kb *Eco*RI fragment of NF-E1c, to create in vitro expression subclones of NF-E1b and NF-E1c. Translation of these two transcripts initiates from the ATG in native sequence context.

NF-E1a and GF-1 subclones were linearized with *Nde*I and transcribed by using T3 RNA polymerase; NF-E1b and NF-E1c subclones were linearized with *Hin*dIII and transcribed with SP6 RNA polymerase. Five micrograms of linearized template was transcribed after the method of Hull et al. (1988), in which

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

a cap analog [$m^7G(5')ppp(5')G$; Pharmacia] was included to subsequently favor efficient translation. An aliquot of RNA from the reactions was denatured and electrophoresed on a 1.5% agarose gel containing 2.2 M formaldehyde and then stained with ethidium bromide. RNA ladders (BRL) were used as standards to determine the size of the transcripts, and 18S and 28S rRNAs present in a known amount of total RNA were used as a standard to determine concentrations.

Rabbit reticulocyte lysate (Promega) was employed to translate 2.5 μg of the capped transcripts in vitro. [^{35}S]Methionine (55 μCi), 100 μg/ml of PMSF, and 100 μg/ml of leupeptin were included in each reaction mixture. For the synthesis of unlabeled proteins (used in the DNA-binding assays), cold methionine was added to the reaction to a final concentration of 20 μM. Two microliters of the translation reaction was electrophoresed on a 15% SDS–polyacrylamide gel (Laemmli 1970) beside ^{14}C-labeled standards (Amersham). The gel was then fixed and subjected to fluorography (Bonner and Laskey 1974).

DNA-binding assays

Two different oligonucleotides were used for the DNA-binding assay (Fried and Crothers 1981; Garner and Revzin 1981): an oligonucleotide from the mouse α1-globin promoter (MαP; Tsai et al. 1989) and one from the chicken β-globin enhancer (CβE; Gallarda et al. 1989), containing either one or two binding sites of the WGATAR motif, respectively. Sense-strand oligonucleotides were end-labeled with T4 polynucleotide kinase and annealed to a fourfold excess of the corresponding unlabeled antisense oligonucleotide. Probe (0.5 ng) was used in each binding reaction.

The binding buffer consisted of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 4% Ficoll, 1 mM DTT, and 75 mM KCl. An equimolar mixture of poly[d(I-C)] and poly[d(A-T)] (Boehringer-Mannheim) was used as nonspecific competitor at concentrations indicated in the figure legends. Fifty- or 100-fold excess of unlabeled CβE or MαP, respectively, was used as sequence-specific competitor. Twofold more MαP oligonucleotide than CβE oligonucleotide was used in these competitions because the former contains only a single binding site, whereas the latter has two sites. Translation mixture (1.5 μl) was added last to each binding reaction. As a positive control, 0.5 μl of basic extract of DNA binding proteins from adult erythroid cells (Gallarda et al. 1989) was allowed to react with the same probes. Binding reactions (15 μl) were incubated for 30 min at room temperature and electrophoresed on 5% nondenaturing polyacrylamide gels in 0.25 × TBE buffer (22 mM Tris-borate, 22 mM boric acid, 0.5 mM EDTA) at 4°C on 18-cm gels at 225 V for 1.5 hr.

Trans-*activation plasmid constructions*

To perform cotransfection *trans*-activation assays, plasmids that constitutively express NF-E1 proteins were constructed by using the vector TFAneo (Federspiel et al. 1989), where the inserted cDNAs are transcriptionally directed by the RSV LTR promoter and enhancer. The cDNAs were first subcloned into an adaptor plasmid, CLA12NCO (Hughes et al. 1987). *Nco*I–*Eco*RI fragments from p33d (NF-E1a; 1.0 kb) and p30a (NF-E1b; 2.4 kb) were subcloned into the *Nco*I–*Eco*RI sites of CLA12NCO. A 1.5-kb *Nco*I/blunt-ended fragment of GF-1 cDNA was inserted into the *Nco*I–*Sma*I sites of CLA12NCO. As a result, these three constructs now have an initiator ATG and untranslated leader sequence, both of which are derived from the v-*src* gene of RSV. The entire NF-E1c cDNA (1.8 kb) was excised by *Xba*I/*Hin*dIII digestion and transferred into CLA12NCO and is therefore translated from its intrinsic initiation codon. The NF-E1 and GF-1 cDNAs were then inserted into the *Cla*I site of TFAneo by using the polylinker array in CLA12NCO, which is flanked by *Cla*I sites. The pMT2 vector-derived *trans*-activation constructs (Fig. 8) were prepared similarly (Kaufman et al. 1989; D.I.K. Martin and S.H. Orkin, in press).

Transfection and cells

Cotransfection activation of reporter genes was carried out as described (Giguere et al. 1986). QT6 cells were trypsinized at confluence, and 40% of the cells were then transfected with a mixture of (genuine or mock) activator plus reporter gene plasmids in addition to an internal transfection control plasmid (RSV luciferase: Wood et al. 1989). Cells were transfected by using $CaPO_4$ (Trainor et al. 1987); after 12 hr in transfection salts at 37°C, 3% CO_2, the plates were washed once with PBS, fed with fresh DME/10% FBS, and transferred to 37°C, 5% CO_2 (zero time).

For each transfection analysis, optimization for activator (constitutive promoter driving the putative factor cDNA under analysis) to reporter gene was determined by varying the ratios of both plasmids and comparing this value to the same ratio of activator plasmid with inserted antisense cDNA as a mock *trans*-activator. The ratios were found to vary greatly: In chicken and mammalian cells, 7 μg of pMT2 activator to 3 μg of reporter was found to give the greatest stimulatory activity, whereas use of the RSV LTR activator (in TFAneo)/reporter in a ratio of 2 : 8 μg, respectively, was found to be optimal.

In analysis of SV1CAT (Gorman et al. 1982) bearing three copies of the CβE-binding site, transfection efficiency was as-

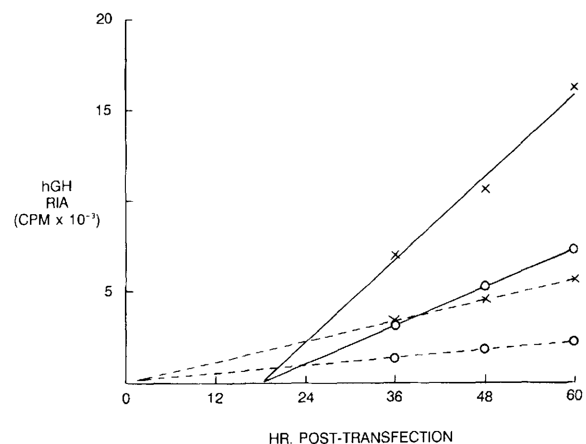


Figure 8. Kinetics of growth hormone secretion. In a preliminary assay, an NF-E1b cDNA clone (transcriptionally directed by the adenovirus major late promoter within plasmid vector pMT2; Kaufman et al. 1989) was cotransfected into QT6 cells with the M6αGH reporter gene at various ratios to optimize for *trans*-activation. To examine whether or not secretion of the reporter hGH peptide was linear with time over the course of a normal experiment, supernatants were withdrawn over a 24-hr period and subsequently assayed by RIA (see Methods). The lines represent hGH secretion in the presence of either 5 μg (×) or 7.5 μg (○) of pMT2/NF-E1b plus 5 or 2.5 μg of M6αGH reporter plasmid, respectively; the dashed lines represent hGH secretion in the presence of equal amounts of vector (without activator cDNA) plus M6αGH at the same concentrations.

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

sessed by normalizing for the activity of a cotransfected RSV–luciferase plasmid (20 ng) cotransfected with the activator and reporter plasmids. Amounts of extract containing identical luciferase activity were then used for the CAT assays. For the reporter gene assays presented (Fig. 7), four plasmids were used: All employed the rabbit β-globin TATA box and transcription initiation site (Fig. 7A; solid box) to direct transcription of the hGH gene (Fig. 7A; open box) in plasmid p0GH (Nichols Institute Diagnostics, San Juan Capistrano, CA; S.H. Orkin, unpubl.). To this base construct, either one or six copies of the MαP oligonucleotide or three or six copies of the CβE oligonucleotide were inserted directly 5′ to the TATA box to produce the M1αGH, M6αGH, C3βGH, and C6βGH plasmids, respectively (Fig. 7A; arrows). After supernatants containing secreted hGH were collected and assayed by RIA (using the Allegro hGH system from Nichols Institute Diagnostics), the cells were scraped from the plates, gently lysed, and examined for DNA binding activity of the proteins expressed from the activator plasmids by gel mobility-shift assay (Tsai et al. 1989; Fig. 7B). In separate experiments, it was determined that *trans*-activation, as measured by the assays used here, is also reflected by a parallel increase in the abundance of correctly initiated activator mRNA (D.I.K. Martin and S.H. Orkin, in press) *Trans*-activation was calculated by the following equation:

$$\textit{Trans}\text{-activation} = \frac{(A + R) - M}{R - M}$$

where $A + R$ represents counts per minute recovered in RIA after transfection with activator plus reporter plasmids, M is counts per minute in the media, and R represents counts per minute recovered after transfection with the reporter alone.

We noted in initial assays that accumulation of the hGH reporter product in cotransfection with the *trans*-activation plasmid was linear (with time) only after ~18 hr post-transfection. When the vector alone (minus the activator cDNA) was included in mock activation with the reporter plasmid, activator protein-independent reporter gene product accumulated linearly beginning at time zero (Fig. 8). We assume that this delay in factor-dependent stimulation of the reporter gene is due to the time required for accumulation of the *trans*-acting factor inside cotransfected cells. To circumvent this repetitive error in recording activation data, medium for all hGH assays was replaced 24 hr after transfection; the assays were then performed on supernatants collected after an additional 24 hr.

Acknowledgments

We thank Z. Yang, K. Lim, and S.J. Stamler for technical assistance during various phases of these studies, K.P. Foley, D.I.K. Martin, and L. Zon for critical evaluation of the manuscript, and C. Hofland and R. Gaber for expert assistance in preparing the figures. We are grateful to S. Hughes and M. Federspeil (Frederick Cancer Research and Development Center) for providing the RSV expression plasmid (TFAneo) used in the *trans*-activation assays, and to B. Morris and R. Mierendorf (Promega and Novagen, Madison, WI) for generous assistance and expertise in the construction of the RBC4 cDNA library. We thank P. Hayman (SUNY Stony Brook) for allowing us to use the v-*rel*(RCAS)-transformed lymphoblasts prior to publication. We also thank N. Hayashi (Tohoku U.) for his encouragement in this work. This work was supported by grants from the National Institutes of Health (to S.H.O. and J.D.E.) and by an NRSA Postdoctoral Fellowship Award (to M.W.L.). S.H.O. is an investigator of the Howard Hughes Medical Institute.

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

References

Akiyama, Y. and S. Kato. 1974. Two cell lines from lymphomas of Marek's disease. *Biken J.* **17:** 105–116.

Behringer, R.R., R.E. Hammer, R.L. Brinster, R.D. Palmiter, and T.M. Townes. 1987. Two 3′ sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. *Proc. Natl. Acad. Sci.* **84:** 7056–7060.

Benton, W.D. and R.W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science* **196:** 180–182.

Beug, H., A. Leutz, P. Kahn, and T. Graf. 1984. ts mutants of E26 leukemia virus allow transformed myeloblasts, but not erythroblasts or fibroblasts to differentiate at the nonpermissive temperature. *Cell* **39:** 579–588.

Beug, H., H. Muller, S. Grieser, G. Doederlein, and T. Graf. 1981. Hemopoietic cells transformed in vitro by REV-T avian reticuloendotheliosis virus express characteristics of very immature lymphoid cells. *Virology* **115:** 295–309.

Beug, H., S. Palmieri, C. Freudenstein, H. Zentgraf, and T. Graf. 1982. Hormone-dependent terminal differentiation in vitro of chicken erythroleukemia cells transformed by ts mutants of avian erythroblastosis virus. *Cell* **28:** 907-919.

Bonner, W.M. and R.A. Laskey. 1974. A film detection method for tritium-labelled proteins and nucleic acids in polyacrylamide gels. *Eur. J. Biochem.* **46:** 83–88.

Bruns, G.A. and V.M. Ingram. 1973. The erythroid cells and hemoglobins of the chick embryo. *Philos. Trans. R. Soc. Lond. Ser. B.* **266:** 225–305.

Catala, F., E. deBoer, G. Habets, and F. Grosveld. 1989. Nuclear protein factors and erythroid transcription of the human $^{A}\gamma$-globin gene. *Nucleic Acids Res.* **17:** 3811–3826.

Chisholm, D. 1989. A convenient moderate-scale procedure for obtaining DNA from bacteriophage lambda. *Biotechniques* **7:** 21–23.

Choi, O.-R. and J.D. Engel. 1986. A 3′ enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. *Nature* **323:** 731–734.

———. 1988. Developmental regulation of β-globin gene switching. *Cell* **55:** 17–26.

Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore, and P.A. Sharp. 1988. The B-cell specific Oct-2 protein contains POU box and homeo box-type domains. *Genes Dev.* **2:** 1570–1581.

Cleveland, D.W., M.A. Lopata, R.J. MacDonald, N.J. Cowan, W.J. Rutter, and M. Kirschner. 1980. Number and evolutionary conservation of α- and β-tubulin and cytoplasmic α- and γ-actin genes using specific cloned cDNA probes. *Cell* **20:** 95–105.

Emerson, B.M., J.M. Nickol, P.D. Jackson, and G. Felsenfeld. 1987. Analysis of the tissue-specific enhancer at the 3′ end of the chicken adult β-globin gene. *Proc. Natl. Acad. Sci.* **84:** 4786–4790.

Emerson, B.M., J.M. Nickol, and T.C. Fong. 1989. Erythroid-specific activation and derepression of the chick β-globin promoter in vitro. *Cell* **57:** 1189–1200.

Enver, T., N. Raich, A.J. Ebens, T. Papayannopoulou, F. Constantini, and G. Stamatoyannopoulos. 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* **344:** 309–313.

Evans, T. and G. Felsenfeld. 1989. The erythroid-specific transcription factor Eryf1: A new finger protein. *Cell* **58:** 877–

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

885.

Evans, T., M. Reitman, and G. Felsenfeld. 1988. An erythrocyte specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc. Natl. Acad. Sci.* **85:** 5976–5980.

Federspiel, M.J., L.B. Crittenden, and S.H. Hughes. 1989. Expression of avian reticuloendotheliosis virus envelope confers host resistance. *Virology* **173:** 167–177.

Forrester, W.C., S. Takegawa, T. Papayannopoulou, G. Stamatoyannopoulos, and M. Groudine. 1987. Evidence for a locus activation region: The formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res.* **15:** 10159–10177.

Fried, M.G. and D.M. Crothers. 1981. Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids. Res.* **9:** 6505–6525.

Gallarda, J.L., K.P. Foley, Z. Yang, and J.D. Engel. 1989. The β-globin stage selector element factor is erythroid-specific promoter/enhancer binding protein NF-E4. *Genes Dev.* **3:** 1845–1859.

Garner, M.M. and A. Revzin. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9:** 3047–3060.

Giguere, V., S.M. Hollenberg, M.G. Rosenfeld, and R.M. Evans. 1986. Functional domains of the human glucocorticoid receptor. *Cell* **46:** 645–652.

Gorman, C.M., L.F. Moffat, and B.H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2:** 1044–1051.

Grosveld, F., G.B. van Assendelft, D.R. Greaves, and G. Kollias. 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51:** 975–985.

Groudine, M., M. Peretz, and H. Weintraub. 1981. Transcriptional regulation of hemoglobin switching in chicken embryos. *Mol. Cell. Biol.* **1:** 281–288.

Gübler, U. and B.J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* **25:** 263–269.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28:** 351–359.

Hesse, J.E., J.M. Nickol, M.R. Lieber and G. Felsenfeld. 1986. Regulated gene expression in transfected primary chicken erythrocytes. *Proc. Natl. Acad. Sci.* **83:** 4312–4316.

Ho, I.C., L.-H. Yang, G. Morle, and J.M. Leiden. 1989. A T-cell-specific transcriptional enhancer element 3′ of Cα in the human T-cell receptor α locus. *Proc. Natl. Acad. Sci.* **86:** 6714–6718.

Hope, I.A. and K. Struhl. 1985. GCN4 protein, synthesized in vitro, binds *his*3 regulatory sequences: Implications for general control of amino acid biosynthetic genes in yeast. *Cell* **43:** 177–189.

Hughes, S.H., J.J. Greenhouse, C.J. Petropoulos, and P. Sutrave. 1987. Adaptor plasmids simplify the insertion of foreign DNA into helper-independent retroviral vectors. *J. Virol.* **61:** 3004–3012.

Hull, J.D., R. Gilmore, and R.A. Lamb. 1988. Integration of a small integral membrane protein, M_2, of influenza virus into the endoplasmic reticulum: Analysis of the internal signal-anchor domain of a protein with an ectoplasmic NH_2 terminus. *J. Cell. Biol.* **106:** 1489–1498.

Johnson, P.F. and S.L. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58:** 799–839.

Kaufman, R.J., M.V. Davies, V.K. Pathak, and J.W.B. Hershey. 1989. The phosphorylation state of eucaryotic initiation factor 2 alters translational efficiency of specific mRNAs. *Mol. Cell Biol.* **9:** 946–958.

Khazaie, K., T.J. Dull, T. Graf, J. Schlessinger, A. Ullrich, H. Beug, and B. Vennström. 1988. Truncation of the human EGF receptor leads to differential transforming potentials in primary avian fibroblasts and erythroblasts. *EMBO J.* **7:** 3061–3071.

Kim, H.-R,C., B.S. Kennedy, and J.D. Engel. 1989. Two chicken erythrocyte band 3 mRNAs are generated by alternative transcriptional initiation and differential RNA splicing. *Mol. Cell. Biol.* **9:** 5198–5206.

Knezetic, J.A. and G. Felsenfeld. 1989. Identification and characterization of a chicken α-globin enhancer. *Mol. Cell. Biol.* **9:** 893–901.

Kollias, G., J. Hurst, E. deBoer, and F. Grosveld. 1987. The human β-globin gene contains a downstream developmental specific enhancer. *Nucleic Acids Res.* **15:** 5739–5747.

Kozak, M. 1989. The scanning model for translation: An update. *J. Cell Biol.* **108:** 229–241.

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

Lewis, C.D., S.P. Clark, G. Felsenfeld, and H. Gould. 1988. An erythrocyte-specific protein that binds to the poly(dG) region of the chicken β-globin gene promoter. *Genes Dev.* **2:** 863–873.

Maniatis, T., S. Goodbourn, and J.A. Fischer. 1987. Regulation of inducible and tissue-specific gene expression. *Science* **236:** 1237–1245.

Martin, D.I.K. and S.H. Orkin. 1990. Transcriptional activation and DNA-binding by the erythroid factor GF-1/NF-E1/Eryf1. *Genes Dev.* **4** (in press).

Martin, D.I.K., S.-F. Tsai, and S.H. Orkin. 1989. Increased γ-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* **338:** 435–438.

Mignotte, V., J.F. Eleouet, N. Raich, and P.-H Romeo. 1989. Cis- and trans-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. *Proc. Natl. Acad. Sci.* **86:** 6548–6552.

Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* **14:** 1319–1324.

Moscovici, C., M.G. Moscovici, H. Jimenez, M.M.C. Lai, M.J. Hayman, and P.K. Vogt. 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. *Cell* **11:** 95–103.

Mueller, C.R., P. Maire, and U. Schibler. 1990. DBP, a liver-enriched transcription activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. *Cell* **61:** 279–291.

Nickol, J.M. and G. Felsenfeld. 1988. Bidirectional control of the chicken β- and ϵ-globin genes by a shared enhancer. *Proc. Natl. Acad. Sci.* **85:** 2548–2552.

Perkins, N.D., R.H. Nicolas, M.A. Plumb, and G.H. Goodwin. 1989. The purification of an erythroid protein which binds to enhancer and promoter elements of haemoglobin genes. *Nucleic Acids Res.* **17:** 1299–1314.

Plumb, M., J. Frampton, H. Wainwright, M. Walker, K. Macleod, G. Goodwin, and P. Harrison. 1989. GATAAG; a cis-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. *Nucleic Acids Res.* **17:** 73–92.

Ptashne, M. 1988. How eukaryotic transcriptional activators

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

work. *Nature* **325:** 683–689.

Redondo, J.M., S. Hata, C. Brocklehurst, and M.S. Krangel. 1990. A T cell-specific transcriptional enhancer within the human T cell receptor δ locus. *Science* **247:** 1225–1229.

Reitman, M., and G. Felsenfeld. 1988. Mutational analysis of the chicken β-globin enhancer reveals two positive-acting domains. *Proc. Natl. Acad. Sci.* **85:** 6267–6271.

Riddle, R.D., M. Yamamoto, and J.D. Engel. 1989. Expression of δ-aminolevulinate synthase in avian cells: Separate genes encode erythroid-specific and nonspecific isozymes. *Proc. Natl. Acad. Sci.* **86:** 792–796.

Rosen, C.A., J.G. Sodroski, and W.A. Haseltine. 1985. The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. *Cell* **41:** 813–823.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Sap, J., A. Muñoz, K. Damm, Y. Goldberg, J. Ghysdael, A. Leutz, H. Beug, and B. Vennström. 1986. The c-*erbA* protein is a high-affinity receptor for thyroid hormone. *Nature* **324:** 635–640.

Scherer, W.F., J.T. Syverton, and G.O. Gey. 1953. Studies on the progagation in vitro of poliomyelitis viruses. *J. Exp. Med.* **97:** 695–715.

Tanaka, M. and W. Herr. 1990. Differential transcriptional activation by Oct-1 and Oct-2: Interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60:** 375–386.

Todaro, G.J. and H. Green. 1963. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* **17:** 299–313.

Trainor, C.D., S.J. Stamler, and J.D. Engel. 1987. Erythroid-specific transcription of the chicken histone H5 gene is directed by a 3′ enhancer. *Nature* **328:** 827–830.

Trainor, C.D., T. Evans, G. Felsenfeld, and M.S. Boguski. 1990. Structure and evolution of a human erythroid transcription factor. *Nature* **343:** 92–96.

Trudel, M., J. Magram, L. Bruckner, and F. Constantini. 1987. Upstream $^{G}\gamma$-globin and downstream β-globin sequences required for stage-specific expression in transgenic mice. *Mol. Cell. Biol.* **7:** 4024–4029.

Tsai, S.-F, D.I.K. Martin, L.I. Zon, A.D. D'Andrea, G.G. Wong, and S.H. Orkin. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339:** 446–451.

Wall, L., E. deBoer, and F. Grosveld. 1988. The human β-globin gene 3′ enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev.* **2:** 1089–1100.

Weintraub, H. and M. Groudine. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* **193:** 848–856.

Wilson, D.B., D.M. Dorfman, and S.H. Orkin. 1990. A non-erythroid GATA-binding protein is required for function of the human preproendothelin-1 promoter in endothelial cells. *Mol. Cell. Biol.* (in press).

Winoto, A. and D. Baltimore. 1989. A novel, inducible and T cell-specific enhancer located at the 3′ end of the T cell receptor α locus. *EMBO J.* **8:** 729–733.

Wood, W.M., M.Y. Kao, D.F. Gordon, and E.C. Ridgway. 1989. Thyroid hormone regulates the mouse thyrotropin β-subunit gene promoter in transfected primary thyrotropes. *J. Biol. Chem.* **264:** 14840–14847.

Zenke, M., P. Kahn, C. Disela, B. Vennström, A. Leutz, K. Keegan, M.J. Hayman, H.-R. Choi, N. Yew, J.D. Engel, and H. Beug. 1988. v-*erbA* specifically suppresses transcription of the avian erythrocyte anion transporter (band 3) gene. *Cell* **52:** 107–119.

Zon, L.I., S.-F. Tsai, S. Burgess, P. Matsudaira, G.A.P. Bruns, and S.H. Orkin. 1990. The major human erythroid DNA-binding protein (GF-1): Primary sequence and localization of the gene to the X chromosome. *Proc. Natl. Acad. Sci.* **87:** 668–672.

Downloaded from genesdev.cshlp.org on August 23, 2022 - Published by Cold Spring Harbor Laboratory Press

Activity and tissue-specific expression of the transcription factor NF-E1 multigene family.

M Yamamoto, L J Ko, M W Leonard, et al.

Genes Dev. 1990, **4:**
Access the most recent version at doi:10.1101/gad.4.10.1650

References This article cites 73 articles, 30 of which can be accessed free at:
http://genesdev.cshlp.org/content/4/10/1650.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

horizon
a PerkinElmer company
Streamline your research with
Horizon Discovery's ASO tool

Copyright © Cold Spring Harbor Laboratory Press