

Induction of Human Fetal Globin Gene Expression by a Novel Erythroid Factor, NF-E4

WENLAI ZHOU,¹ DAVID R. CLOUSTON,¹ XI WANG,² LORETTA CERRUTI,¹
JOHN M. CUNNINGHAM,² AND STEPHEN M. JANE^{1*}

*Rotary Bone Marrow Research Laboratory, Royal Melbourne Hospital Research Foundation, Parkville, Victoria, Australia,¹
and Division of Experimental Hematology, St. Jude Children's Research Hospital, Memphis, Tennessee²*

Received 23 May 2000/Returned for modification 21 June 2000/Accepted 18 July 2000

The stage selector protein (SSP) is a heteromeric complex involved in preferential expression of the human γ -globin genes in fetal-erythroid cells. We have previously identified the ubiquitous transcription factor CP2 as a component of this complex. Using the protein dimerization domain of CP2 in a yeast two-hybrid screen, we have cloned a novel gene, NF-E4, encoding the tissue-restricted component of the SSP. NF-E4 and CP2 coimmunoprecipitate from extract derived from a fetal-erythroid cell line, and antiserum to NF-E4 ablates binding of the SSP to the γ promoter. NF-E4 is expressed in fetal liver, cord blood, and bone marrow and in the K562 and HEL cell lines, which constitutively express the fetal globin genes. Enforced expression of NF-E4 in K562 cells and primary erythroid progenitors induces endogenous fetal globin gene expression, suggesting a possible strategy for therapeutic intervention in the hemoglobinopathies.

The human β -globin cluster is the classic paradigm of a multigene locus. The globin genes (ϵ , $\zeta\gamma$, $\alpha\gamma$, δ , and β) are expressed at high levels throughout ontogeny in a stringently regulated developmental stage- and tissue-specific pattern. From conception until week 5 of gestation, the embryonic globin (ϵ) gene is expressed in the yolk sac, the major site of erythropoiesis. After this time, the first switch in globin subtype occurs, as the fetal globin ($\zeta\gamma$ and $\alpha\gamma$) genes become the dominant transcripts in the erythropoietic cells of the fetal liver. This expression pattern persists until birth, when the switch from fetal to adult globin (β -globin) synthesis occurs, coincident with the bone marrow becoming the predominant erythropoietic organ (26, 43, 57).

Studies of the human β -globin locus in transgenic mice and in patients carrying the Hispanic and Dutch thalassemic mutations have revealed that the key regulatory sequences required for high-level globin expression reside 6 to 20 kb upstream of the ϵ gene (15, 31, 61, 64, 65). These sequences, characterized by the presence of five 5' DNase I hypersensitivity sites (5'HS1 to -5), are known as the locus control region (LCR) (16, 20, 62). Studies by Wijgerde et al. suggest that the HSs act cooperatively as a holocomplex which focuses the vast enhancing potential of the LCR to a single globin gene at any given time point during ontogeny (67). In murine fetal liver cells transgenic for the β -globin locus, the LCR flip-flops back and forth between the γ and β genes at the time of the fetal-adult switch. As the cellular transcription factor milieu changes to favor adult globin expression, the stability of the γ gene-LCR interaction decreases and β -globin becomes the predominantly transcribed gene.

Competition between globin genes for a single regulatory sequence was first proposed as a mechanism of developmental regulation by Choi and Engel (8). In these studies, a stage selector element (SSE) in the chick β -gene (β) promoter was essential for the preferential interaction of that promoter with

the locus enhancer during adult erythropoiesis. The activity of the promoter element was mediated through the binding of a stage-specific factor, known as NF-E4 (19, 69). Promoter sequences and stage-specific factors have also been shown to be critical for correct developmental regulation of the human and murine β -globin clusters. Mice carrying a transgene of the human locus lacking the LCR, or embryonic stem cells in which the native LCR has been removed by homologous recombination, still display appropriate temporal patterns of globin expression, albeit at reduced levels (13, 58). Conversely, deletion of the human γ promoter or its substitution with a non-developmental-stage-specific erythroid promoter in transgenic mice abolishes the correct temporal profile of both γ - and β -gene expression (3, 53).

Two regions of the γ promoter appear to be responsible for its competitive advantage in the fetal-erythroid environment. The first, the CACCC box, binds a recently described member of the Krüppel family, fetal Krüppel-like factor (FKLF) (5). Expression of this gene is detectable in fetal liver and to a lesser extent adult bone marrow, but its functional effects appear to predominantly involve the ϵ and γ genes. The second region in the γ promoter was defined in transfection studies in the K562 cell line, a model of fetal erythropoiesis (37). In these experiments, an 18-bp stage selector element immediately 5' of the TATA box was sufficient for preferential transcription from the γ promoter when in competition with the β promoter for a single enhancer element (HS2) from the LCR (28). Analogous to the chicken cluster, the activity of this SSE was dependent on the binding of a stage-specific factor, the fetal-erythroid cell-specific stage selector protein (SSP) (29).

Several lines of evidence support the importance of the SSP-SSE interaction in the developmental regulation of globin gene expression. Evolutionary phylogenetic footprinting studies demonstrate absolute conservation of the SSP binding site in species with a distinct stage of fetal globin expression of the γ genes and loss in species where the γ genes are embryonic (21, 60). Multiple SSP binding sites have also been identified in phylogenetic footprints in the ϵ promoter, HS2 and HS3 (22). The formation of a new binding site for the SSP by the -202(C→G) HPFH mutation also lends credence to a role for this factor in γ -gene regulation (27). Biochemical purification

* Corresponding author. Mailing address: Rotary Bone Marrow Research Laboratory, Royal Melbourne Hospital Research Foundation, c/o Royal Melbourne Hospital Post Office, Grattan St., Parkville, VIC 3050, Australia. Phone: 61-3-93428641. Fax: 61-3-93428634. E-mail: jane@wehi.edu.au.

of the SSP revealed that the ubiquitously expressed transcription factor CP2 (also known as LBP-1c/LSF) formed a major component of the SSP binding activity (29, 30, 35, 70). Antiserum to CP2 ablated the SSP-SSE complex in electrophoretic mobility shift assays (EMSA). It also reacted with highly purified chicken NF-E4 in Western analysis, indicating that this developmental complex is conserved in evolution (69). However, CP2 alone was incapable of binding to the SSE, suggesting that the SSP consisted of a heteromeric complex between CP2 and an unknown factor that provided the tissue specificity and DNA binding activity of the complex. The presence of this factor (named NF-E4, after the putative chick homologue) was confirmed in EMSA and UV cross-linking experiments, and its molecular mass was estimated as 40 to 45 kDa (29).

Using the protein dimerization domain of CP2 as bait in a yeast two-hybrid library screen, we now report the cloning and characterization of human NF-E4, the tissue-restricted component of the SSP. We demonstrate that enforced expression of this factor in fetal-erythroid cells induces fetal and embryonic globin expression.

MATERIALS AND METHODS

Yeast two-hybrid screen. The cDNA sequence encoding the COOH-terminal 240 amino acids (aa 260 to 502) of CP2 was inserted into the yeast expression vector pGBT9 (Clontech). The resultant plasmid encodes a hybrid protein containing the DNA binding domain of GAL4 (GAL4DBD) fused in frame to CP2 residues. The yeast reporter strain, HF7C, rendered competent by the lithium acetate method, was sequentially transformed with this vector and a plasmid cDNA library derived from K562 cells constructed in the yeast expression vector pACT2 (14). The cDNAs in this vector were fused with the GAL4 transactivation domain (GAL4AD). The yeast cells were plated on leucine-, tryptophan-, and histidine-deficient plates and incubated at 30°C for 4 days. Potential protein interactions were indicated by activation of the histidine reporter gene and growth on these plates and by activation of the second reporter gene, β -galactosidase, and a positive 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) assay. Library plasmids were rescued from yeast clones using the acid-washed glass beads procedure and electroporated into the competent *Escherichia coli* strain MC1061.

Sequencing reactions were performed using the Taq Dyedeoxy terminator cycle sequencing method (Applied Biosystems) and analyzed on an Applied Biosystems 373 automated sequencer.

Yeast one-hybrid assay. A concatemer of four copies of the SSE was cloned into the *EcoRI/SalI* sites of the yeast vector, pLacZ. As a control, a four-copy concatemer of the direct repeat elements (DREs) from the proximal β promoter was also cloned into this vector. Small-scale transformations of each vector were performed into *Saccharomyces cerevisiae* YM4271, which is auxotrophic for histidine, uracil, leucine, and tryptophan. Prior to transfection, the vector was linearized with *NotI* to allow genomic integration into the *ura3-52* site, which confers auxotrophy to uracil, allowing selection of transformants. A single yeast colony which displayed no basal reporter gene activity was chosen for subsequent experiments. This colony was expanded and transformed with pACT-CP2, pACT106, or pACT117. The transformants were selected on minimal medium lacking leucine and uracil, and colonies were lifted on filters and assayed for β -galactosidase activity.

Mammalian two-hybrid assay. Mammalian expression vectors containing the dimerization domain of CP2 fused in frame to the GAL4DBD and NF-E4 fused in frame to the VP16AD were generated. These plasmids were cotransfected with pG5CAT, a reporter construct with five GAL4DBDs linked to the chloramphenicol acetyltransferase (CAT) gene, into 293 cells, using calcium phosphate precipitation. Vectors lacking either CP2 or NF-E4 were transfected as controls. After 48 h, cells were harvested and whole-cell lysate was prepared. CAT activity was measured by CAT enzyme-linked immunosorbent assay as instructed by the manufacturer (Boehringer Mannheim).

5' RACE. A marathon 5' RACE (rapid amplification of 5' cDNA ends) cDNA library was constructed from mRNA from K562 cells as instructed by the manufacturer (Clontech). Nested PCR was performed with the following vector- and gene-specific primers: gene specific 1 (5'-CCCTTGGCTCAGATGAAGCGATGGTAGT-3'), gene specific 2 (5'-TGGCCTGCAGGGCCCCAGTAGGT-3'), vector specific 1 (5'-CCATCCTAATACGACTACTATAGGGC-3'), and vector specific 2 (5'-ACTACTATAGGGCTCGAGCGGC-3'). PCR conditions were as follows: 95°C for 1 min, 1 cycle; 94°C for 10 s and 68°C for 2 min, 30 cycles; and 68°C for 5 min, 1 cycle. Nested PCR was performed under identical conditions except that the cycle number was reduced to 20. PCR products were electrophoresed on 1% agarose, blotted onto nitrocellulose, and probed with internal gene-specific oligonucleotides. Final PCR products were cloned into the TOPO 2.1 vector as instructed by the manufacturer (Invitrogen) and sequenced.

Generation of MSCV-based supernatant and transduction of mammalian cell lines. The NF-E4 coding region was cloned into the retroviral vector plasmid MSCV-HA at a unique *XhoI* or *EcoRI* site. This bicistronic vector contains (i) the amphotropic retrovirus murine stem cell virus (MSCV) 5' long terminal repeat (LTR), (ii) a hemagglutinin (HA) epitope tag with the NF-E4 coding sequence in frame either 5' or 3' to the tag, (iii) the encephalomyocarditis internal ribosomal entry site (IRES), (iv) the green fluorescent protein (GFP) cDNA, and (v) the MSCV 3' LTR (see Fig. 2C). The plasmid was cotransfected with an amphotropic packaging plasmid into 293T cells by calcium phosphate precipitation. After 48 h, the supernatant containing amphotropic particles was harvested, filtered, and added to K562 or MEL cells every 12 h for 3 days. The cells were allowed to recover for 72 h and then analyzed for GFP expression by flow cytometry. The highest-expressing 10% of cells were sterilely sorted, expanded, resorted, and subsequently expanded in oligoclonal pools. A biological titer of the supernatant on NIH 3T3 cells was equivalent to 10^6 CFU/ml.

Generation of RD18 producer cell lines. Amphotropic supernatant generated in 293T cells (as described above) was used to transfect the FLYRD18 packaging cell line (10). Briefly, fresh filtered supernatant from 293T cells was added to RD18 cells plated at a density of 10^3 cells every 12 h for 3 days. Subsequently, the top 20% of GFP-positive RD18 cells were obtained by fluorescence-activated cell sorting (FACS) and cultured until confluent. Amphotropic supernatant harvested from these plates was used to transfect CD34⁺ progenitors. The expression of NF-E4 in the producer cell line was verified by immunoblotting with anti-HA antiserum (data not shown).

Isolation and retroviral transduction of human CD34⁺ cells. Human cord blood was generously provided by the Bone Marrow Donor Institute Cord Blood Bank. CD34⁺ cells were isolated using a MiniMACS magnetic cell sorting system (Miltenyi Biotec Inc.). Cells were then cultured overnight with expansion medium, which contains 1% deionized bovine serum albumin (BSA; Stem Cell Technologies), insulin (5 μ g/ml; Sigma), transferrin (100 μ g/ml; BRL), low-density lipoprotein (10 μ g/ml; Sigma), 10^{-4} M β -mercaptoethanol (BRL), recombinant human interleukin-3 (rhIL-3; 10 ng/ml; R&D), rhIL-6 (10 ng/ml; R&D), recombinant human stem cell factor (300 ng/ml; R&D), and Flt-3 (300 ng/ml; R&D). Non-tissue culture-treated 35-mm-diameter dishes were coated with RetroNectin CH286 solution (TaKaRa Biochemicals, Shiga, Japan) at the concentration of 20 μ g/cm² for 2 h at room temperature and then blocked with 2% BSA fraction V (Fisher Scientific) for 30 min at room temperature (40). The coated dishes were preloaded with virus supernatant from RD18 producer lines (2 ml/well) for 30 min, after which the supernatant was removed. Another 1 ml of supernatant was added along with expanded CD34⁺ cells at $<5 \times 10^5$ cells/dish; then 1 ml of expanded medium (in which the amount of each ingredient was doubled) was added and mixed well. Cells were cultured for 24 h and then harvested and carefully washed three times with a large volume of 1 \times phosphate-buffered saline. The cells were then cultured at 10^5 cells/ml in Iscove modified differentiation medium containing 30% fetal calf serum (HyClone), 1% deionized BSA (Stem Cell Technologies), recombinant human stem cell factor (100 ng/ml; R&D), rhIL-3 (0.1 pg/ml; R&D), human erythropoietin (10 U/ml; Amgen), insulin (10 μ g/ml; Sigma), 10^{-4} M β -mercaptoethanol (BRL), 10^{-6} M hydrocortisone (Stem Cell Technologies), and penicillin-streptomycin and glutamine (1:1,000; BRL) as previously described (59). At day 12, GFP- and glyco-phorin A-positive cells were isolated by FACS, and RNA was prepared using RNazol (Tel-Test Inc.).

Expression of GST fusion proteins and affinity chromatography. CP2 and NF-E4 cDNAs were cloned in frame with the glutathione S-transferase (GST) coding sequence in the pGEX vectors (Pharmacia). The GST fusion proteins were expressed in *E. coli* BL21. Fusion proteins were purified on glutathione-Sepharose (Pharmacia), and their integrity was confirmed with Coomassie blue staining after sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). For in vitro protein-protein interaction assays, 1 μ g of GST or GST fusion protein was incubated for 1 h at 4°C with 10 μ l of glutathione-Sepharose beads, which had been preblocked with 0.5% milk. After extensive washing, the beads were resuspended in 200 μ l of binding buffer (10 mM Tris-HCl [pH 7.9], 500 mM KCl, 0.1 mM EDTA, 150 μ g of BSA/ml, 0.1% Nonidet P-40, 10% glycerol) and incubated for 1 h at room temperature with [³⁵S]methionine-labeled NF-E4. After extensive washing, retained proteins were eluted by boiling in SDS loading buffer and analyzed by SDS-PAGE and autoradiography.

Extract preparation, immunoprecipitation, and EMSA. Nuclear extracts were prepared by the method of Dignam (11). Highly purified SSP was obtained by fractionating crude extract over heparin-Sepharose and DNA affinity columns as described previously (29).

For immunoprecipitation studies, nuclear extracts were initially precleared with normal rabbit serum (10 μ g/ml) and then incubated with preimmune serum or antiserum to CP2 or NF-E4 overnight at 4°C. A 50% slurry of protein G-Sepharose was added and incubated at 4°C for 1 h. The mixture was then centrifuged at 3,000 \times g for 1 min, and the pellet was washed in 50 mM Tris-HCl (pH 7.9) containing 150 mM NaCl prior to being resuspended in SDS loading buffer. Samples were subjected to SDS-PAGE, transferred to polyvinylidene difluoride (PVDF) membranes, and blotted with antiserum to NF-E4. Signal detection was achieved with the Amersham Pharmacia ECL (enhanced chemiluminescence) system according to the manufacturer's instructions. EMSAs were performed by incubating various amounts of nuclear extract with 10^5 cpm of [³²P]dCTP-end-labeled double-stranded oligonucleotides encoding the SSE re-

gion of the γ promoter in a 20- μ l reaction containing 500 ng of poly(dI-dC), 6 mM MgCl₂, 16.5 mM KCl, and 100 μ g of BSA. For antibody studies, 3 μ l of preimmune serum or rabbit anti-mouse CP2 or NF-E4 antibody was preincubated for 10 min with the binding reaction prior to addition of the probe. After incubation on ice for 15 min and 25°C for 15 min, samples were electrophoresed on a 4% nondenaturing polyacrylamide gel in 0.5 \times Tris-borate-EDTA buffer for 90 min at 10 V/cm.

RNase protection, reverse transcription-PCR (RT-PCR), and Northern analysis. RNase protection analysis was performed using an Ambion RNase protection assay kit according to the manufacturer's instructions. Probes used in these studies were as described previously (40a). Probe input was 10⁶ cpm/sample for γ - and β -globin probes and 0.25 \times 10⁶ cpm/sample for the 18S probe.

For RT-PCR, first-strand cDNA was prepared from 2 μ g of mRNA from primary tissues using random hexamers. Each cDNA sample was appropriately diluted to give similar amplification of S14 RNA under the same PCR conditions. The primer sequences and PCR conditions were as follows: S14 sense (5'-GGCAGACCGAGATGAATCCTCA-3') and antisense (5'-CAGGTCCAGGGGTC TTGGTCC-3'), 95°C for 1 min, followed by various cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 5 min; NF-E4 sense (5'-ACCCGGGAGGGGCTCCGGTCTT-3') and antisense (5'-CCCTTGGCT CAGATGAAGCGATGGTAGT-3'), 95°C for 1 min, followed by various cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 5 min. All PCR products were electrophoresed on 1.5% agarose gels, transferred to nitrocellulose, and analyzed by Southern blotting using ³²P-radiolabeled internal oligonucleotides as probes. Membranes were then autoradiographed for 2 h at -70°C. Northern analysis of K562 pools was performed as described previously (54).

RESULTS

Isolation of CP2-interacting proteins from a K562 cell cDNA library. Previous studies had demonstrated that the ubiquitous transcription factor CP2 formed a major component of the SSP. We and others have mapped the protein dimerization domain of CP2 to the 242 amino acid residues at its carboxy terminus (56, 63). Within this region, a 17-aa stretch (aa 292 to 309) is essential for protein-protein interactions (S. M. Jane and J. M. Cunningham, unpublished data). A cDNA sequence encoding the COOH-terminal 242 aa (aa 260 to 502) of CP2 was inserted into the yeast expression vector pGBT9. The resultant plasmid (GAL4CP2-260) encodes a hybrid protein containing the GAL4DBD fused to CP2 residues 260 to 502 (Fig. 1A). The yeast reporter strain HF7C was transformed with this vector and an expression library derived from K562 cell line cDNAs fused to the sequences encoding the GAL4AD. This library was chosen because K562, a human cell line, is a model of fetal erythropoiesis, constitutively expressing the ϵ and γ genes but not the adult β genes (51). In addition, abundant SSP binding activity is evident in nuclear extract from these cells, which was the source for the biochemical purification of the CP2 component of the SSP. From 5 \times 10⁶ clones screened, we isolated 100 clones that appeared to interact with the CP2 bait. From this collection we identified about 40 clones that encoded CP2, an expected result in view of the protein's ability to homodimerize. Another 50 clones represented previously identified false positives from the two-hybrid screen. Of the 10 remaining clones, 8 corresponded to known genes or ESTs (expressed sequence tags) whose tissue distribution suggested that they were unlikely to represent the tissue-restricted component of the SSP. Only two clones, c106 and c117, were novel, and hence their further evaluation was prioritized.

As the DNA binding activity of the SSP appears to depend on the presence of the CP2 partner protein, we evaluated the ability of c106 and c117 to bind to the SSE in the yeast one-hybrid assay. Yeast strain YM4271, containing four concatenated SSE sites linked to a *lacZ* reporter gene, was generated and transfected with the K562 cDNA library plasmids encoding the GAL4AD-c106, GAL4AD-c117, or GAL4AD-CP2 fusion protein. A strain containing four concatenated DREs from the proximal β promoter linked to the same re-

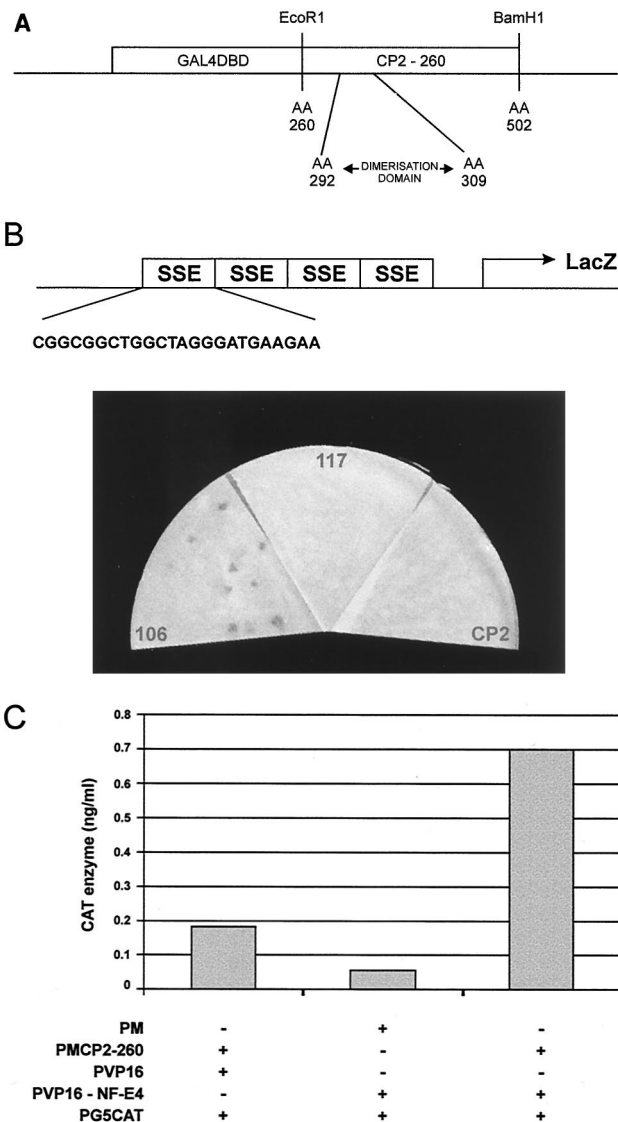


FIG. 1. Isolation of CP2-interacting proteins. (A) Schematic of the bait construct used in the yeast two-hybrid screen of a K562 cDNA library. The coding sequence of CP2 from aa 260 to 502 was fused in frame with the GAL4DBD. The minimal dimerization domain (aa 292 to 309) is shown. (B) Yeast one-hybrid assay of CP2-interacting clones defined in the two-hybrid screen. Plasmids containing CP2 or c106 or c117 fused to the GAL4AD were transfected into a yeast reporter strain carrying four concatenated SSE sites linked to a *lacZ* reporter gene. The sequence of a single SSE site is expanded. Transformants were selected on minimal medium plates lacking leucine and uracil, lifted onto filters, and assayed for β -galactosidase activity. Filters were photographed after 30 min, and equivalent transfection efficiencies were observed for all constructs. (C) Mammalian two-hybrid analysis of interaction between CP2 and NF-E4. 293 cells were transiently cotransfected with a reporter plasmid containing five GAL4DBD sites linked to the CAT gene (pG5CAT) and the indicated expression vectors. PM contains the GAL4DBD, PMCP2-260 contains the GAL4DBD fused to aa 260 to 502 of CP2, PVP16 has the VP16AD, and PVP16-NF-E4 has the NF-E4 sequence isolated from the two-hybrid clone fused to the VP16AD. Transfections were performed in triplicate, and cell lysates were assayed in triplicate for CAT activity.

porter was transfected as a control. After transfection, both strains were plated on medium lacking leucine and uracil, and resultant colonies were assayed for β -galactosidase activity. As seen in Fig. 1B, a positive result was observed with c106 but not with c117 or CP2 using the SSE binding site. No enzymatic

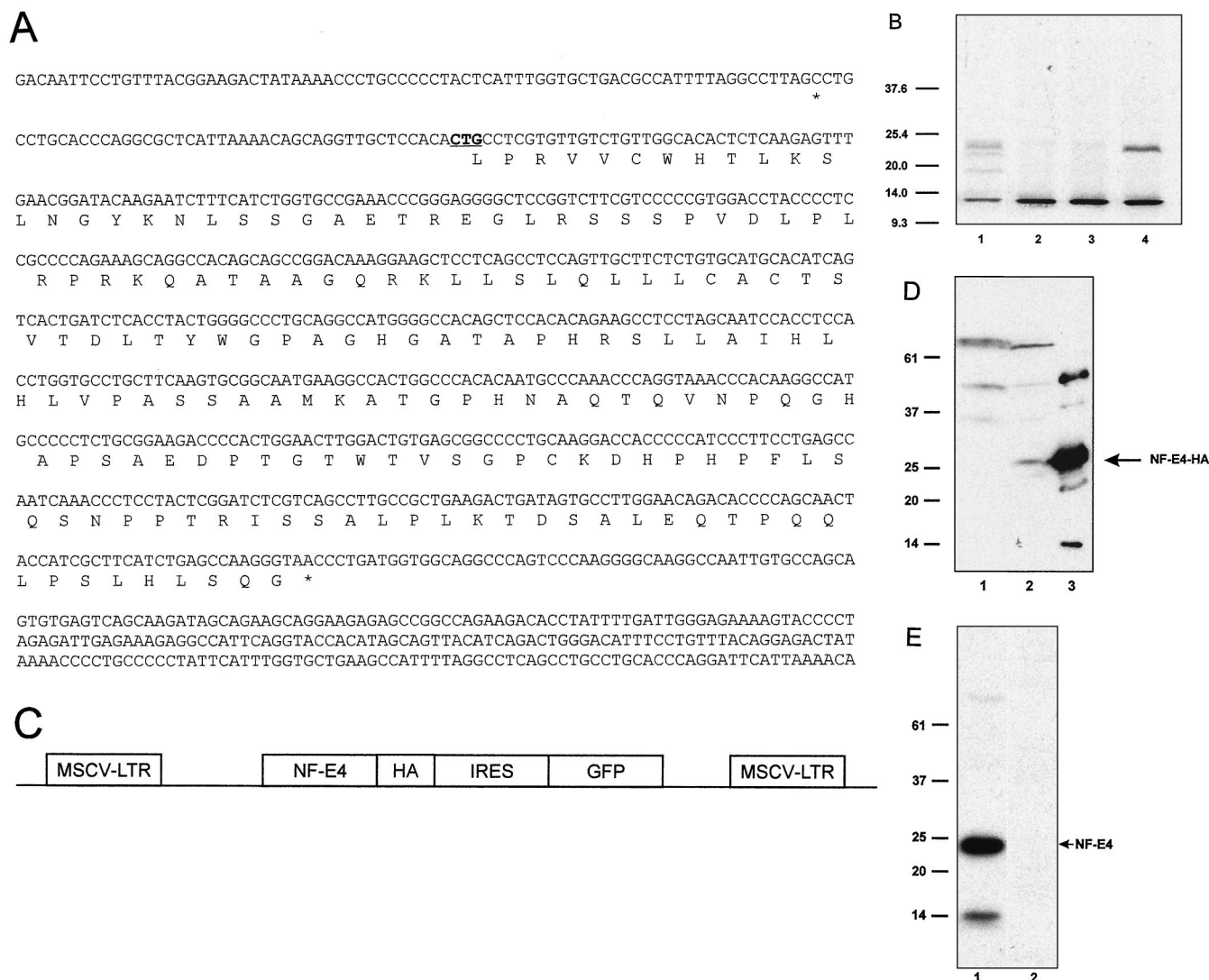


FIG. 2. NF-E4 is a 22-kDa protein which may initiate at a CUG codon. (A) Nucleotide and amino acid sequences of NF-E4. The potential initiator CUG is underlined and in boldface. The 5' and 3' termination codons are marked with asterisks. (B) In vitro transcription/translation of NF-E4. The NF-E4 cDNA with its native Kozak sequence was cloned into pSP72, and [³⁵S]methionine-labeled protein was produced (lane 1). We also generated radiolabeled protein from NF-E4 constructs in which the CTG was mutated to a GGG or TTG but retained the native Kozak sequence (lanes 2 and 3) or was mutated to an ATG with a consensus Kozak sequence (lane 4); 2 μ l of each sample was resolved by SDS-PAGE on a 12% gel. The gel was dried and subjected to autoradiography. The migration of molecular size standards is marked in kilodaltons. (C) Diagrammatic representation of the MSCV-NF-E4-HA vector. The vector consists of the MSCV backbone containing the NF-E4 coding sequence tagged at the COOH terminus with HA, followed by an encephalomyocarditis virus IRES linked to the GFP cDNA. (D) Western analysis of K562 cell pools expressing NF-E4-HA. Whole-cell lysates from K562 cells transduced with MSCV-NF-E4-HA (lane 2) or MSCV alone (lane 1) were resolved on a 12% polyacrylamide gel, transferred to a PVDF membrane, probed with polyclonal anti-HA antiserum, and developed with ECL. The specific NF-E4-HA band is arrowed. Recombinant NF-E4-HA served as a control (lane 3). The migration of molecular mass standards is indicated in kilodaltons. (E) Western analysis of native K562 cells. Nuclear extract from K562 cells was resolved on a 12% polyacrylamide gel, transferred to a PVDF membrane, and probed with polyclonal anti-NF-E4 (lane 1) or preimmune (lane 2) serum. Signal was developed with ECL. The migration of molecular mass standards is indicated in kilodaltons.

activity was observed with the DRE binding sites with any of the three plasmids (data not shown). Based on these findings, we postulated that c106 was a strong candidate for the partner protein of CP2 in the SSP complex and hereafter refer to it as NF-E4.

To validate the interaction between CP2 and NF-E4 in a eukaryotic expression system, we used the mammalian two-hybrid assay. Mammalian expression vectors containing the dimerization domain of CP2 fused in frame to the GAL4DBD and NF-E4 fused in frame to the VP16AD were generated. These plasmids were cotransfected with pG5CAT, a reporter construct with five GAL4 DBDs linked to the CAT gene.

Vectors lacking either CP2 or NF-E4 were transfected as controls. As seen in Fig. 1C, a marked induction of CAT activity was observed only in the presence of both proteins.

The NF-E4 gene encodes a 22-kDa protein which may initiate at a CUG codon. To facilitate further studies of NF-E4, we used 5' RACE from K562 cell cDNA to obtain a full-length clone. A 966-bp fragment was generated using nested gene- and vector-specific primers in multiple experiments. Comparison of the sequences derived from multiple clones with the databases using the BLAST algorithm revealed a high degree of homology with a sequence from a bacterial artificial chromosome containing a region of the human X chromosome.

Sequence analysis revealed a long open reading frame contiguous with that defined in the original yeast two-hybrid GAL4AD-c106 fusion vector (Fig. 2A). Although one potential initiation codon (AUG) was observed beginning at nucleotide 421, several observations suggested that translation of full-length NF-E4 might not start at this AUG. First, the NF-E4 reading frame remains open for an additional 115 codons upstream of the first AUG before an in-frame termination codon is encountered. Second, the predicted size of the protein from the first AUG is markedly less than that suggested by our previous studies. Finally, a CUG codon preceded by a Kozak sequence and termination codon is present in the correct reading frame 100 codons upstream of the first AUG (32). Translation from this codon would generate a protein with a predicted molecular mass of approximately 22 kDa.

Many studies have documented translation of human mRNAs from initiation codons other than AUG, most often CUG (33, 38, 46). These include transcription factors such as TEF-1 and Krox-24 and proto-oncogene products such as c-myc, Int2, and others (1, 23, 34, 68). Comparison of the NF-E4 sequence with a human genomic clone isolated in our laboratory confirmed the presence of the single AUG in the midregion of the sequence and the upstream CUG and termination codon (data not shown). To determine whether effective translation was achieved from this CUG *in vitro*, we subcloned the NF-E4 cDNA into the *in vitro* transcription/translation vector pSP72 and generated [³⁵S]methionine-labeled protein. We also generated NF-E4 clones in which the CTG was mutated to a GGG or TTG or to an ATG with a consensus Kozak sequence to determine whether these sequence changes resulted in altered efficiency in translation (32). As shown in Fig. 2B, the dominant protein species translated from the AUG containing NF-E4 vector (lane 4) had a molecular weight identical to that observed with the native CUG-initiated NF-E4 (lane 1). A marked reduction in efficiency of transcription and translation was observed with both the GGG- and TTG-containing constructs (lanes 2 and 3).

To determine whether effective translation was achieved from this CUG *in vivo*, we generated an MSCV-based vector containing the NF-E4 cDNA (24). This bicistronic vector contains the GFP cDNA linked by the encephalomyelitis virus IRES to the NF-E4 cDNA tagged at its COOH terminus with HA (Fig. 2C). K562 cells were transduced with this virus (MSCV-NF-E4-HA) or the parental virus carrying the GFP cDNA alone (MSCV); after 5 days, GFP-positive cells were selected by FACS. As shown in Fig. 2D, Western analysis with an anti-HA antibody demonstrated a band in extract from the MSCV-NF-E4-HA-transduced cells (lane 2) which comigrated with recombinant NF-E4-HA generated in bacteria (lane 3). No corresponding band was observed in K562 cells transduced with MSCV alone (lane 1).

Further support for CUG-initiated translation was obtained with the generation of polyclonal antiserum to the native NF-E4 protein. A dominant band of 22 kDa was observed in Western analysis of native K562 cells, consistent with initiation at the CUG codon (Fig. 2E). A second minor band was observed at approximately 14 kDa, which could reflect initiation at the downstream AUG.

NF-E4 interacts with CP2 *in vitro* and *in vivo*. To confirm the interaction between CP2 and full-length NF-E4, we used GST chromatographic assays. GST, GST fused in frame with full-length endophilin (GST-END), or GST fused in frame with full-length CP2 (GST-CP2) was coupled to glutathione-Sepharose beads and incubated under stringent conditions with [³⁵S]methionine-labeled *in vitro*-transcribed/translated NF-E4. Specific retention of NF-E4 was observed with the GST-CP2 beads but not control GST or GST-END beads (Fig.

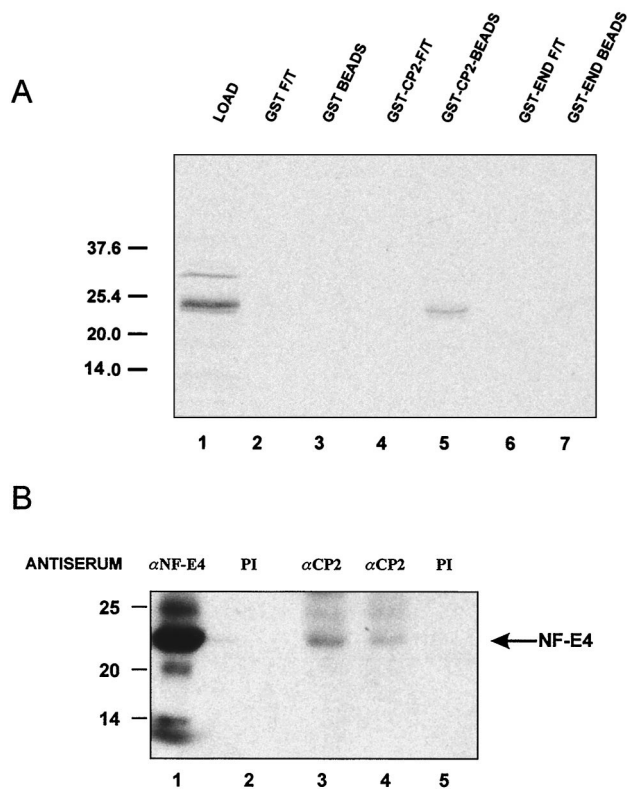


FIG. 3. CP2 and NF-E4 interact *in vitro* and *in vivo*. (A) Direct interaction between CP2 and NF-E4. GST-CP2, GST-END, or GST alone was expressed in *E. coli*, and 1 μ g of protein bound to glutathione-Sepharose beads. The beads were then incubated with 2 μ l of [³⁵S]-labeled *in vitro*-translated NF-E4 in binding buffer for 1 h at room temperature (see Materials and Methods). After extensive washing, the beads were resuspended in SDS loading buffer and subjected to SDS-PAGE (lanes 1, 3, and 5). The supernatants (flowthrough [F/T]) of the binding reactions were also analyzed (lanes 2, 4, and 6). Positions of migration of the NF-E4 load and the molecular weight standards are labeled in kilodaltons. (B) Coimmunoprecipitation of CP2 and NF-E4 from native K562 cells. Nuclear extract from K562 cells was immunoprecipitated with polyclonal antiserum to NF-E4 (lane 1) or CP2 (lanes 3 [8 μ l] and 4 [4 μ l]) or the corresponding preimmune (PI) serum (lane 2 or 5, respectively). The immunoprecipitates were fractionated by SDS-PAGE and blotted with antisera to NF-E4. Positions of migration of NF-E4 and the molecular mass standards (in kilodaltons) are labeled.

3A). To confirm this interaction in an *in vivo* setting, coimmunoprecipitation studies were performed. Nuclear extract from K562 cells was immunoprecipitated with either anti-CP2 antiserum or preimmune serum and blotted with anti-NF-E4 antiserum (Fig. 3B). Immunoprecipitation and blotting with anti-NF-E4 antiserum served as the positive control (lane 1). A specific band of 22 kDa was observed after immunoprecipitation with 8 μ l (lane 3) or 4 μ l (lane 4) of anti-CP2 antiserum. No band was observed with preimmune serum derived from NF-E4- or CP2-inoculated rabbits (lanes 2 and 5). This finding indicates that CP2 and NF-E4 form a physiological complex *in vivo*.

NF-E4 is a component of the SSP complex. To confirm that NF-E4 contributed to the formation of the SSP, we examined the effect of anti-NF-E4 antiserum on the SSP-SSE interaction in an EMSA. As shown in Fig. 4A, addition of either anti-CP2 (lane 2) or anti-NF-E4 (lane 4) antiserum to crude K562 cell nuclear extract specifically ablated the formation of the SSP-SSE complex, leaving the Sp1-SSE complex unaltered. Addition of preimmune sera had no effect (lanes 1 and 3).

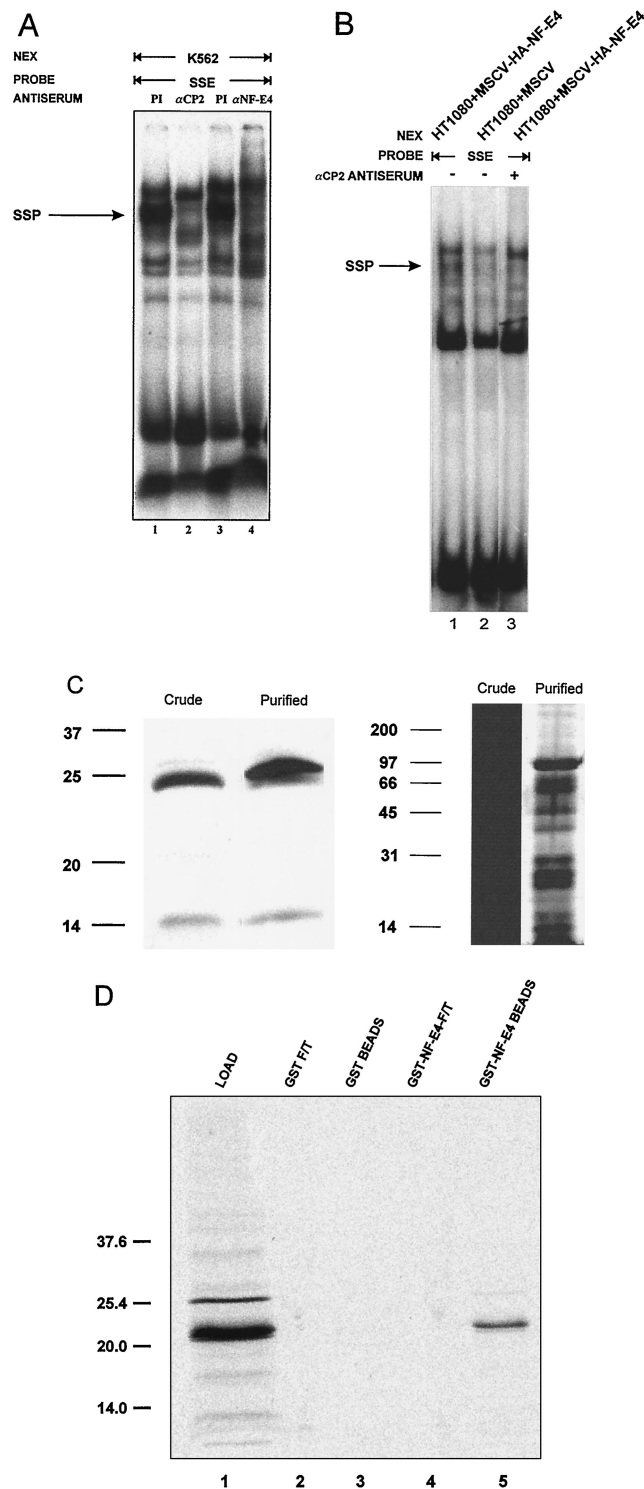


FIG. 4. NF-E4 is a component of the SSP. (A) NF-E4 antiserum disrupts the SSP-SSE complex. Crude K562 cell nuclear extract (NEX) was studied in EMSA with the SSE probe in the presence of anti-CP2 antiserum (lane 2), anti-NF-E4 antiserum (lane 4), or the corresponding preimmune (PI) serum (lane 1 or 3, respectively). Migration of the SSP-SSE complex is arrowed. (B) Expression of NF-E4 in a null cell line generates a de novo SSP complex. Nuclear extract was prepared from the human sarcoma cell line HT1080 transduced with MSCV-HA-NF-E4 (lane 1) or MSCV (lane 2) and analyzed by EMSA using an SSE probe. The effect of anti-CP2 antisera on the EMSA from MSCV-HA-NF-E4 was examined in lane 3. Migration of the SSP complex is arrowed. (C) Western analysis and silver stain of NF-E4 in purified SSP. Crude K562 cell nuclear extract or purified SSP was resolved on a 12% polyacrylamide gel, transferred

To determine whether de novo expression of NF-E4 in a CP2-expressing NF-E4 null cell line could result in the formation of the SSP complex, we transduced the human sarcoma cell line HT1080 with an MSCV-based retrovirus containing the NF-E4 cDNA tagged at the NH₂ terminus with HA (MSCV-HA-NF-E4). This cell line contains abundant CP2 but no NF-E4 at the RNA or protein level (data not shown). HT1080 cells transduced with MSCV alone served as the control. GFP-positive cells were selected by FACS, and expression of NF-E4 was confirmed by Western analysis (data not shown). EMSA using an SSE probe revealed the presence of a new complex in nuclear extract derived from MSCV-HA-NF-E4-transduced cells which comigrated with native SSP (Fig. 4B, lane 1). No complex was observed in the line transduced with MSCV alone (lane 2). Addition of anti-CP2 antisera to this EMSA ablated the SSP-SSE complex (lane 3). Further evidence for the role of NF-E4 in formation of the SSP complex was derived from Western analysis of an SSP fraction purified by heparin-Sepharose and DNA affinity chromatography. As shown in the left panel of Fig. 4C, both the 22-kDa and the lower-molecular-mass NF-E4 species were detected in crude and purified samples.

One confounding result from our previous UV cross-linking studies was the prediction that the molecular mass of the CP2 partner protein in the SSP was 40 to 45 kDa (29). In view of this, we examined whether the 22-kDa NF-E4 protein could form homodimeric complexes in GST chromatography experiments. GST alone or fused in frame with full-length NF-E4 (GST-NF-E4) was coupled to glutathione-Sepharose beads and incubated under stringent conditions with [³⁵S]methionine-labeled in vitro-transcribed-translated NF-E4. Specific retention of NF-E4 was observed with the GST-NF-E4 beads (lane 3) but not control GST beads (lane 1) (Fig. 4D). This finding coupled with our UV cross-linking data suggests that the SSP complex is composed of two NF-E4 molecules linked to a molecule of CP2.

NF-E4 demonstrates a highly restricted pattern of expression. To determine the tissue distribution of NF-E4 expression, we initially performed Northern analysis on mRNA derived from K562 cells. Despite the fact that we had demonstrated the presence in these cells of NF-E4 mRNA (by RT-PCR) and protein (by Western blotting and EMSA), we were unable to detect a signal with a variety of NF-E4 cDNA probes (data not shown). Additional analysis of multitissue Northern blots also failed to detect a signal in a variety of tissues, as did RNase protection analysis on mRNA from tissues and cell lines (data not shown). We therefore proceeded to define the expression pattern of NF-E4 using RT-PCR. Based on the genomic sequence, we designed primers from exons 1 and 2 which are separated by an 1,800-bp intron. The identity of the correct-size PCR product was confirmed by Southern analysis using an internal primer as a probe. As shown in Fig. 5A, NF-E4 is expressed in fetal liver, cord blood, and bone marrow. No

to a PVDF membrane, probed with polyclonal anti-NF-E4 antiserum and developed with ECL (left); the samples were also electrophoresed, and the gel was silver stained (right). Molecular size standards are indicated in kilodaltons. (D) NF-E4 forms homodimeric complexes. GST-NF-E4 or GST alone was expressed in *E. coli*, and 1 μg of protein was bound to glutathione-Sepharose beads. The beads were then incubated with 2 μl of ³⁵S-labeled in vitro-translated NF-E4 in binding buffer for 1 h at room temperature (see Materials and Methods). After extensive washing, the beads were resuspended in SDS loading buffer and subjected to SDS-PAGE (lanes 1 and 3). The supernatants (flowthrough [F/T]) of the binding reactions were also analyzed (lanes 2 and 4). Positions of migration of NF-E4 and the molecular mass standards (in kilodaltons) are labeled.

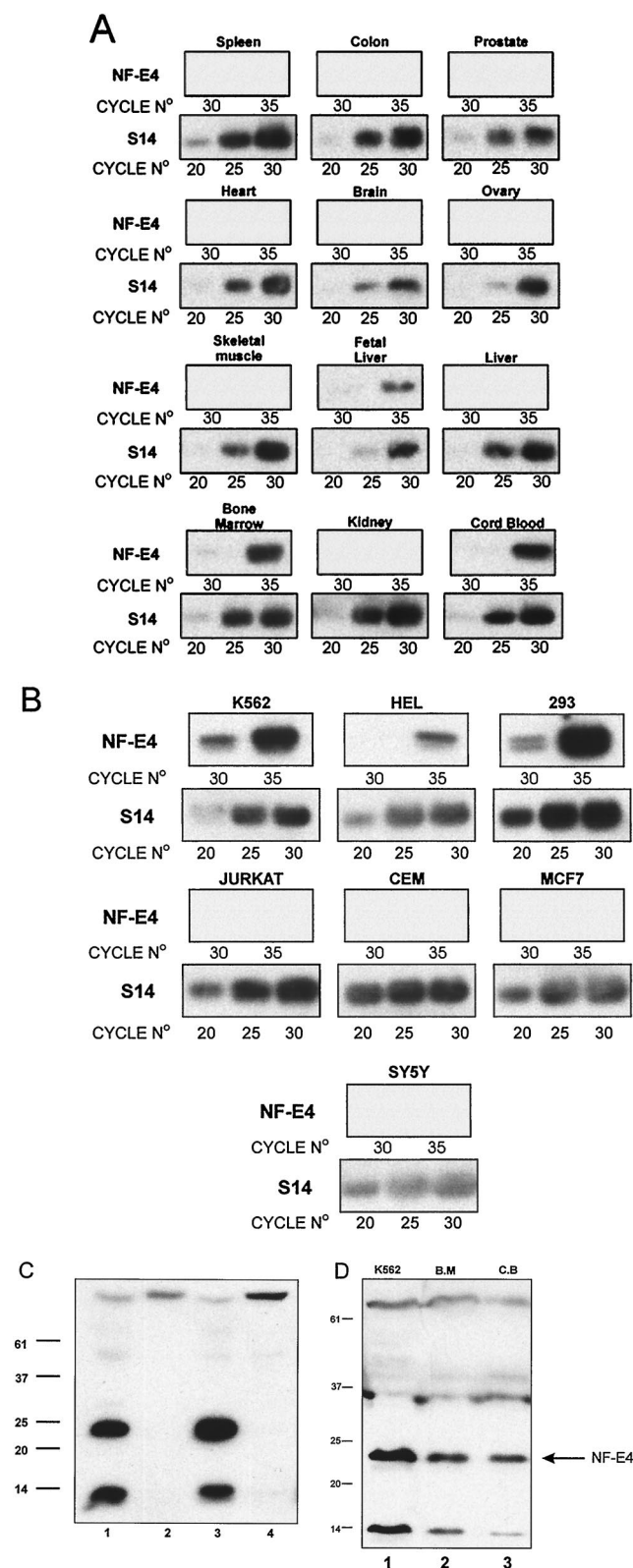


FIG. 5. Expression of NF-E4 in primary tissues and established cell lines. First-strand cDNA transcribed from poly(A)⁺ RNA from multiple primary tissues or cell lines was used as a template to PCR amplify a product using primers specific for S14. Samples were then diluted and reamplified to give comparable band intensities for the same number of amplification cycles of S14 RNA within the linear range of the assay; this represents 20, 25, and 30 cycles. (A) Expression of NF-E4 in primary human tissues. Based on the S14 quantitation, comparable

expression was observed from a variety of other organs, including colon, heart, spleen, kidney, liver, lymph node, and thymus (Fig. 5A and data not shown). In RT-PCR analysis of cell lines, expression was demonstrated with mRNA derived from the fetal and erythroid cell lines K562 and HEL and the embryonic kidney cell line 293T. No product was amplifiable from a variety of other lines, including Jurkat, CEM, MCF7, DU528, SY5Y, and COS (Fig. 5B and data not shown). To confirm the expression of NF-E4 at the protein level, nuclear extract from 293T, COS, K562, and HeLa cell lines was analyzed by Western blotting. As shown in Fig. 5C, the previously defined immunoreactive species were detected in K562 (lane 3) and 293T (lane 1) extracts. Additional Western analysis (Fig. 5D) of cord blood (lane 3)- and bone marrow (lane 2)-derived nuclear extract demonstrated a dominant band of 22 kDa which comigrated with that observed in control K562 extract (lane 1). In addition, the previously recognized smaller species was also detected in all lanes.

Enforced expression of NF-E4 in K562 cells induces fetal and embryonic globin gene expression. To examine the functional role of NF-E4 in globin gene expression, we used MSCV-HA-NF-E4. K562 cells were transduced with this vector or the parent GFP-containing vector (MSCV) and then sorted twice for green fluorescence by FACS and expanded in oligoclonal pools. All pools were subsequently shown to contain more than 99% GFP-positive cells by FACS analysis (data not shown). Northern analysis of pools derived from MSCV-HA-NF-E4-transduced cells showed a significant upregulation (5- to 10-fold as an average of all pools) of γ -gene expression compared to pools from the MSCV-transduced cells (Fig. 6A). Expression of the housekeeping gene (*GAPDH*) was unchanged between pools.

A previous report has suggested a potential role for the SSP in embryonic gene regulation (22). We therefore examined the effects of enforced expression of NF-E4 on ϵ -gene expression. Total RNA from oligoclonal pools of K562 cells transduced with either the MSCV (lanes 1 to 4) or MSCV-HA-NF-E4 (lanes 5 to 9) was analyzed by Northern blotting (Fig. 6B). Results comparable to those observed for γ -gene expression were obtained, with significant induction of ϵ -gene expression in the cells transduced with MSCV-HA-NF-E4. No evidence of β -gene activation was observed in any pool (data not shown).

To evaluate the specificity of NF-E4 activity, we transduced

amounts of cDNA from multiple primary human tissues were PCR amplified using primers specific for NF-E4. These primers span a 1.8-kb intron and thus discriminate between mRNA and genomic DNA-derived signal. Cycle numbers were chosen to represent the linear range of amplification; this represents 30 and 35 cycles. (B) Expression of NF-E4 in cell lines. Based on the S14 quantitation, comparable amounts of cDNA from multiple established cell lines were PCR amplified using primers specific for NF-E4. RNA was prepared from erythroid lines (K562 and HEL), T-cell lines (CEM and Jurkat), a human embryonic kidney cell line (293), a breast cell line (MCF7), and a brain cell line (SY5Y). All PCR products were electrophoresed on 1% agarose, transferred to nitrocellulose, and probed with an internal radiolabeled oligonucleotide specific for the predicted product. (C) Western analysis of NF-E4 in cell lines. Nuclear extract isolated from 293T (lane 1), COS (lane 2), K562 (lane 3), and HeLa (lane 4) cell lines was resolved on a 12% polyacrylamide gel, transferred to a PVDF membrane, probed with polyclonal anti-NF-E4 antiserum, and developed with ECL. Molecular size standards are indicated in kilodaltons. (D) NF-E4 protein in cord blood and bone marrow. CD34⁺ cells were isolated from fresh bone marrow (B.M; lane 2) or cord blood (C.B; lane 3) and cultured for 7 days (see Materials and Methods). Nuclear extracts were prepared from both samples and resolved by SDS-PAGE using a 12% gel. Crude K562 nuclear extract served as the control (lane 1). After transfer to PVDF, the samples were immunoblotted with anti-NF-E4 antiserum and developed with ECL. Positions of migration of NF-E4 and the molecular size standards (in kilodaltons) are indicated.

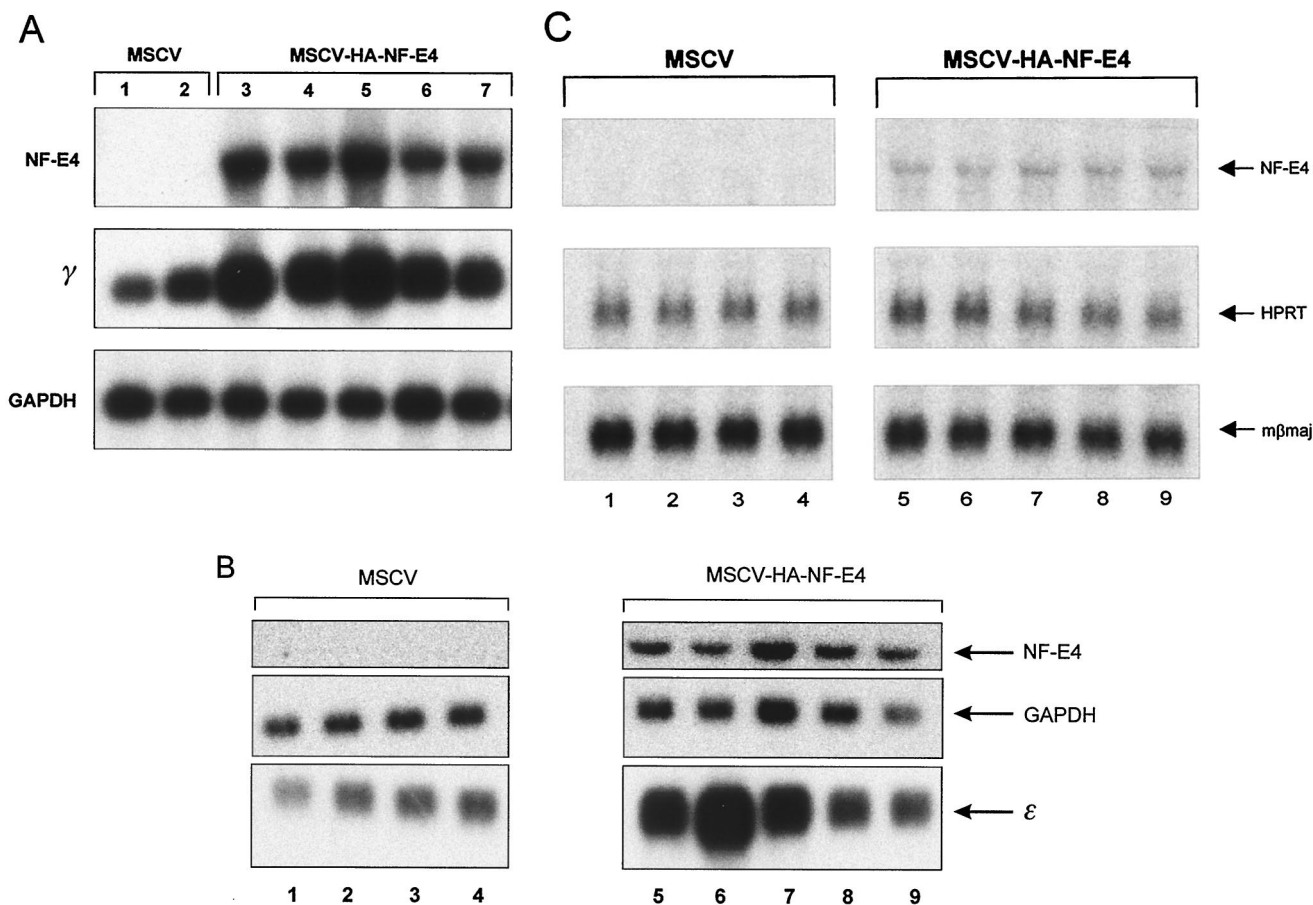


FIG. 6. Enforced expression of NF-E4 induces γ - and ϵ -gene expression. (A) Northern analysis of K562 cell pools overexpressing NF-E4. K562 cells were transduced with either MSCV-HA-NF-E4 or MSCV alone, and GFP-positive cells were obtained by FACS. Cells were expanded, resorted, and cultured in oligoclonal pools; 10 μ g of total RNA from five MSCV-HA-NF-E4 pools (lanes 3 to 7) or two MSCV pools (lanes 1 and 2) was analyzed with γ -gene and NF-E4 probes. *GAPDH* served as the control. (B) Enforced expression of NF-E4 induces ϵ -gene expression. K562 cells were transduced with either MSCV-HA-NF-E4 or MSCV, and GFP-positive cells were obtained by FACS. Cells were expanded, resorted, and cultured in oligoclonal pools; 10 μ g of total RNA from four MSCV pools (lanes 1 to 4) or five MSCV-HA-NF-E4 pools (lanes 5 to 9) was analyzed with ϵ -gene and NF-E4 probes. *GAPDH* served as the control. (C) Northern analysis of MEL cell pools overexpressing NF-E4. MEL cells were transduced with either MSCV-HA-NF-E4 or MSCV alone, and GFP-positive cells were obtained by FACS. Cells were expanded, resorted, and cultured in oligoclonal pools; 10 μ g of total RNA from four MSCV pools (lanes 1 to 4) or five MSCV-HA-NF-E4 pools (lanes 5 to 9) was analyzed with β maj gene and NF-E4 probes. *HPRT* served as the control.

a murine erythroleukemia cell line (MEL) with either MSCV (lanes 1 to 4) or MSCV-HA-NF-E4 (lanes 5 to 9) and sorted for GFP expression. Oligoclonal pools were analyzed by Northern blotting. As shown in Fig. 6C, no change in murine β maj expression was observed. Expression of murine ϵ and β H1 was undetectable in all clones (data not shown).

Enforced expression of NF-E4 in cord blood progenitors induces γ -gene and represses β -gene expression. To extend the functional studies of NF-E4, we generated stable FLYRD18 producer cell lines containing either MSCV or MSCV-HA-NF-E4 (10). Supernatants from these lines were used to transduce CD34⁺ cells derived from human cord blood (see Materials and Methods). The cells were then cultured for 12 days in differentiation medium, and GFP- and glycophorin A-positive cells were separated by FACS. RNA was prepared from these cells and analyzed by RNase protection assay. As shown in Fig. 7A, the most striking difference between the MSCV-NF-E4 and control MSCV pools was the reduction in β -gene expression in the NF-E4-transduced pools. After normalization for the housekeeping gene 18S, the reduction in β -gene expression induced by NF-E4 was approximately 10-fold. As the signal

detected with the γ -gene probe in this assay was intense in both MSCV and NF-E4 pools, we examined various dilutions of RNA to determine whether a difference was evident at lower concentrations. As shown in Fig. 7B, analysis in the linear portion of the assay revealed a twofold increase in γ -gene expression in pools transduced with MSCV-NF-E4 compared with those transduced with the control MSCV retrovirus.

DISCUSSION

In this report we detail the molecular cloning and characterization of human NF-E4, a novel gene encoding the tissue-restricted component of the SSP complex. The gene was isolated from a yeast two-hybrid screen of a K562 cell cDNA library using CP2, the previously identified ubiquitous component of the SSP, as the bait. As predicted from our previous studies (29), NF-E4 is essential for DNA binding of the SSP, as demonstrated by the disruption of the SSP-SSE complex induced by NF-E4 antiserum and the activation in the yeast one-hybrid assay induced by NF-E4. Based on our GST chromatographic assays and previous UV cross-linking data, it ap-

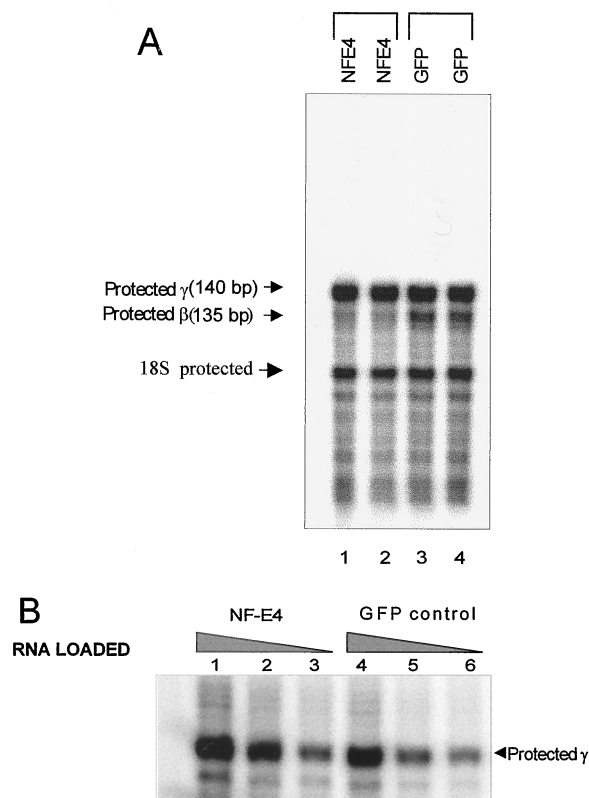


FIG. 7. Enforced expression of NF-E4 in cord blood progenitors induces γ -gene and represses β -gene expression. (A) RNase protection assays on RNA from transduced cord blood CD34⁺ cells. CD34⁺ cells were isolated from fresh cord blood and transduced with either MSCV-HA-NF-E4 or the MSCV generated in the FLYRD18 packaging cell line (see Materials and Methods). Cells were expanded in *in vitro* differentiation culture (59) for 12 days, and then GFP- and glycoprotein A-positive cells were obtained by FACS; 1 μ g of total RNA was used in each lane, and the predicted protected fragments for human γ - and β -globin are shown on the left. (B) Quantitation of γ -gene expression by RNase protection assay. Total RNA (lanes 1 and 4, 0.8 μ g; lanes 2 and 5, 0.4 μ g; lanes 3 and 6, 0.1 μ g) from MSCV-HA-NF-E4 (lanes 1 to 3) or MSCV (lanes 4 to 6) transduced cord blood CD34⁺ cells (as detailed above) was used; the predicted protected fragment for human γ -globin and 18S is shown on the right.

pears that the SSP is composed of two molecules of NF-E4 linked to a single molecule of CP2.

Analysis of the NF-E4 cDNA and protein sequence revealed no homology to known genes or ESTs. Specifically, no known DNA binding, protein dimerization, or transactivation domains were evident. Protein translation appears to commence at a CUG codon. This is supported by our *in vitro* transcription-translation assays, analysis of *in vivo* translation with retroviral vectors, and the molecular weight of the native protein. It is, however, conceivable that a very small region of additional 5' coding sequence exists, despite the multiple 5' RACE clones which terminated at the same nucleotide. Efforts to formally map the 5' end of the transcript by RNase protection and primer extension were unsuccessful due to the rarity of the transcript in expressing cells (data not shown). Despite this, the NF-E4 clone that we have isolated appears to fulfill its predicted functional role.

Non-AUG initiation has been previously reported for a variety of mammalian proteins, most often from CUG codons (6, 33, 46). However, studies of some mammalian proteins and of viral mRNA in mammalian cells have also demonstrated initiation mediated by GUG, ACG, AUA, and AUU (38). Al-

though our evidence for CUG initiation of native NF-E4 is indirect, it is significant that none of these alternate initiating codons are present either 5' or within 125 bp 3' of the CUG. It is also significant that, like many of the other factors with non-AUG-initiated isoforms, NF-E4 may also exist in a truncated form initiated from a downstream AUG (7, 44). This is evidenced by the smaller species observed in Western analyses and immunoprecipitation experiments with anti-NF-E4 antisera. Although it is possible that this smaller species represents a proteolytic cleavage product of NF-E4, the generation of a comigrating protein from an MSCV carrying the NF-E4 cDNA truncated to this ATG suggests that this species represents the product of alternate translation initiation (data not shown). The use of non-AUG initiation codons in many proteins plays a key regulatory role (7, 23, 34, 39, 41, 50, 55, 68). For example, the CUG- and AUG-initiated isoforms of the steroid receptor binding protein Bag-1 and the proto-oncogene products Int2 and Hck-1 differ in their subcellular localization (1, 36, 44). In addition, the CUG isoform of Bag-1 interacts with different protein partners and consequently has a unique functional role (17). Interestingly, the smaller NF-E4 peptide was immunoprecipitated with anti-NF-E4 antisera but not coimmunoprecipitated with anti-CP2 antisera. This suggests that there may be intrinsic functional differences between these species.

The demonstration of NF-E4 mRNA and/or protein in fetal liver, bone marrow, and cord blood raises the question of the developmental stage specificity of the SSP complex. This finding is analogous to the expression pattern observed for another stage-specific globin regulatory factor, EKLF, which is present at both mRNA and protein level in yolk sac, fetal liver, and adult bone marrow (12; Jane and Cunningham, unpublished). Despite this, β -gene expression is observed only in definitive erythroid cells, and mice nullizygous for EKLF demonstrate no abnormalities in primitive erythropoiesis (42, 48). The mechanism underlying this selectivity remains unknown, although the increase in γ -gene transcription observed in the fetal livers of EKLF^{-/-} mice transgenic for the human globin locus suggests that it may be influenced by promoter competition (47, 66). Functionally, NF-E4 also displays a high degree of selectivity, as it induces fetal and embryonic but not adult globin gene expression. The lack of β -gene induction is observed in the context of K562 cells, in which constitutive β -gene expression is absent, and MEL cells, in which high levels of β -globin gene expression are observed. In contrast, induction of fetal and embryonic globin is observed only in cells in which these genes are normally transcribed. This suggests that the chromatin modifications associated with fetal-embryonic gene silencing cannot be altered simply by changing the transcription factor milieu. This observation is comparable to the effects of enforced expression of EKLF, which fails to induce β -gene expression in cells in which the adult globin gene is constitutively silent (12).

The suppression of β -gene and augmentation of γ -gene expression observed in primary cord blood progenitors are of interest, as this is a cell population in which γ - and β -globin are normally expressed concurrently. This finding suggests that in the context of promoter competition, the presence of NF-E4 is sufficient to alter the balance between the promoters, favoring transcription of the fetal gene. This finding is reminiscent of the studies of EKLF nullizygous mice carrying the human β -globin locus yeast artificial chromosome (β -YAC mice), in which enhanced γ -gene expression is observed in the fetal liver due to the diminution of effective competition from the inactive β promoter (47, 66). The difference in magnitude between β -globin suppression and γ -gene activation in the cord blood progenitors in the setting of enforced expression of NF-E4 is

intriguing. One interpretation is that although the γ promoter has the competitive advantage over the β promoter in the presence of NF-E4, allowing its preferential interaction with the LCR, other factors necessary for optimal γ -gene expression are diminishing as the switch from fetal to adult globin expression progresses. This interpretation is consistent with our previous findings which suggest that binding of the SSP confers only weak transcriptional activation in the setting of promoter competition (28). Studies of the effects of enforced expression of NF-E4 in earlier developmental stages in the β -YAC transgenic line will further address this issue (18).

As low levels of fetal globin expression are detectable in adult bone marrow, the presence of some NF-E4 in this tissue is not surprising. However, the levels observed appear similar to those observed in cord blood, in which the expression of the γ genes is appreciably greater. This finding may indicate that another fetal factor such as FKLf is also required for the developmental specificity of γ -gene expression (5).

Our previous studies have suggested that the enhancer activity of the LCR is influenced by the developmentally specific factors EKLf and the SSP (2). This effect is not mediated by direct protein-protein interactions between the stage-specific promoter-bound factors and factors bound to the LCR (19). An alternate model is that EKLf and the SSP alter promoter structure, rendering it more amenable to the effects of the LCR enhanceosome. This model is supported by the loss of HS formation in the β promoter in EKLf nullizygous mice and by the recent demonstration of complex formation between EKLf and the SWI-SNF chromatin remodeling factors (4, 66). Studies addressing the role of NF-E4 in chromatin modification are in progress.

The ultimate goal of defining factors which activate fetal globin is their potential for therapeutic intervention in the hemoglobinopathies. Patients with β -thalassemia and sickle cell disease who also inherit genetic mutations which prolong fetal globin expression after birth have a significantly ameliorated clinical course (49). To date, three factors which augment γ -gene expression in cell lines have been identified. The first of these, the helix-loop-helix protein Id2, is ubiquitously expressed and likely to have pleiotropic effects on gene regulation (26). However, NF-E4 and FKLf have highly restricted patterns of expression and offer promise for both pharmacological manipulation and gene therapy. The ability of enforced expression of NF-E4 in cord blood progenitors to suppress β -gene expression may have significant implications in genetic therapy of sickle cell disease with the dual beneficial effects of enhanced fetal globin expression and reduction of β^S synthesis. Studies of NF-E4 in mouse models of human hemoglobin switching (18) and hemoglobinopathies (9, 45, 52) represent the next step in the evaluation of this factor as a therapeutic tool.

ACKNOWLEDGMENTS

The first two authors contributed equally to this work.

We thank Stuart Orkin, Art Nienhuis, Jerry Adams, and Glenn Begley for critical reading of the manuscript, Robert Hawley for the gift of the MSCV plasmid, Patrick Kelly and Derek Persons for the gift of the RD18 packaging cell line, and Amy McEwan and Helen Zogos for technical assistance. We also thank members of the Jane and Cunningham laboratories for helpful discussions and Art Nienhuis for continuing support.

This work was supported by the NHMRC of Australia, The Wellcome Trust (S.M.J.), the Anti-Cancer Council of Victoria (D.R.C.), NIH PO1 HL53749-03, Cancer Center Support CORE grant P30 CA 21765, the American Lebanese Syrian Associated Charities (ALSAC), and the Assisi Foundation of Memphis.

REFERENCES

- Acland, P., M. Dixon, G. Peters, and C. Dickson. 1990. The subcellular fate of the Int-2 oncoprotein is determined by choice of initiation codon. *Nature* **343**:662-665.
- Amrolia, P. J., J. M. Cunningham, and S. M. Jane. 1998. Maximal activity of an erythroid-specific enhancer requires the presence of specific protein binding sites in linked promoters. *J. Biol. Chem.* **273**:13593-13598.
- Anderson, K. P., J. A. Lloyd, E. Ponce, S. C. Crable, J. C. Neumann, and J. B. Lingrel. 1993. Regulated expression of the human β -globin gene in transgenic mice requires an upstream globin or nonglobin promoter. *Mol. Biol. Cell* **4**:1077-1085.
- Armstrong, J. A., J. J. Bieker, and B. M. Emerson. 1998. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLf in vitro. *Cell* **95**:93-104.
- Asano, H., X. S. Li, and G. Stamatoyannopoulos. 1999. FKLf, a novel Krüppel-like factor that activates human embryonic and fetal β -like globin genes. *Mol. Cell. Biol.* **19**:3571-3579.
- Boeck, R., and D. Kolakofsky. 1994. Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. *EMBO J.* **13**:3608-3617.
- Bruening, W., and J. Pelletier. 1996. A non-AUG translational initiation event generates novel WT1 isoforms. *J. Biol. Chem.* **271**:8646-8654.
- Choi, O. R. B., and J. D. Engel. 1988. Developmental regulation of β -globin gene switching. *Cell* **55**:17-26.
- Ciavatta, D. J., T. M. Ryan, S. C. Farmer, and T. M. Townes. 1995. Mouse model of human β^0 -thalassemia: targeted deletion of the mouse β maj and β min-globin genes in embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **92**:9259-9263.
- Cosset, F.-L., Y. Takeuchi, J.-L. Battini, R. A. Weiss, and M. K. L. Collins. 1995. High-titer packaging cells producing recombinant retroviruses resistant to human serum. *J. Virol.* **69**:7430-7436.
- Dignam, J. D. 1990. Preparation of extracts from higher eukaryotes. *Methods Enzymol.* **182**:194-203.
- Donze, D., T. M. Townes, and J. J. Bieker. 1995. Role of erythroid Krüppel-like factor in human γ - to β -globin switching. *J. Biol. Chem.* **270**:1955-1959.
- Epner, E., A. Reik, D. Cimbara, A. Telling, M. A. Bender, S. Fiering, T. Enver, D. I. K. Martin, M. Kennedy, G. Keller, and M. Groudine. 1998. The β -globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse β -globin locus. *Mol. Cell* **2**:447-455.
- Fields, S., and O. Song. 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**:245-246.
- Forrester, W. C., E. Epner, M. C. Driscoll, T. Enver, M. Brice, T. Papayanopoulou, and M. Groudine. 1990. A deletion of the human β -globin locus activation region causes a major alteration in chromatin structure and replication across the entire β -globin locus. *Genes Dev.* **4**:1637-1649.
- Forrester, W. C., C. Thompson, J. T. Elder, and M. Groudine. 1986. A developmentally stable chromatin structure in the human β -globin gene cluster. *Proc. Natl. Acad. Sci. USA* **83**:1359-1363.
- Froesch, B. A., S. Takayama, and J. C. Reed. 1998. BAG-1L protein enhances androgen receptor function. *J. Biol. Chem.* **273**:11660-11666.
- Gaensler, K. M. L., M. Kitamura, and Y. W. Kan. 1993. Germ-line transmission and developmental regulation of a 150-kb yeast artificial chromosome containing the human β -globin locus in transgenic mice. *Proc. Natl. Acad. Sci. USA* **90**:11381-11385.
- 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.
- 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.
- Gumucio, D. L., K. L. Blanchard-McQuate, H. Heilstedt-Williamson, D. A. Tagle, T. A. Gray, S. A. Tarle, L. Gragowski, M. Goodman, J. L. Slightom, and F. S. Collins. 1991. Gamma-globin gene regulation: evolutionary approaches, p. 277-289. *In* G. Stamatoyannopoulos and A. W. Nienhuis (ed.), *The regulation of hemoglobin switching. Proceedings of the Seventh Conference on Hemoglobin Switching*. The Johns Hopkins University Press, Baltimore, Md.
- Gumucio, D. L., D. A. Shelton, W. J. Bailey, J. L. Slightom, and M. Goodman. 1993. Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the ϵ -globin gene. *Proc. Natl. Acad. Sci. USA* **90**:6018-6022.
- Hann, S. R., K. Sloan-Brown, and G. D. Spotts. 1992. Translational activation of the non-AUG-initiated c-myc 1 protein at high cell densities due to methionine deprivation. *Genes Dev.* **6**:1229-1240.
- Hawley, R. G., F. H. L. Lieu, A. Z. C. Fong, and T. S. Hawley. 1994. Versatile retroviral vectors for potential use in gene therapy. *Gene Ther.* **1**:136-138.
- Holmes, M. L., J. D. Haley, L. Cerruti, W.-L. Zhou, H. Zogos, D. E. Smith, J. M. Cunningham, and S. M. Jane. 1999. Identification of Id2 as a globin regulatory protein by representational difference analysis of K562 cells induced to express γ -globin with a fungal compound. *Mol. Cell. Biol.* **19**:4182-4190.

26. Jane, S. M., and J. M. Cunningham. 1998. Understanding fetal globin gene expression: a step towards effective HbF reactivation in hemoglobinopathies. *Br. J. Hematol.* **102**:415–422.
27. Jane, S. M., D. L. Gumucio, P. A. Ney, J. M. Cunningham, and A. W. Nienhuis. 1993. Methylation enhanced binding of Sp1 to the stage selector element of the human γ -globin gene promoter may regulate developmental specificity of expression. *Mol. Cell. Biol.* **13**:3272–3281.
28. Jane, S. M., P. A. Ney, E. F. Vanin, D. L. Gumucio, and A. W. Nienhuis. 1992. Identification of a stage selector element in the human γ -globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the β -promoter. *EMBO J.* **11**:2961–2969.
29. Jane, S. M., A. W. Nienhuis, and J. M. Cunningham. 1995. Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein. *EMBO J.* **14**:97–105.
30. Kim, C. H., C. Heath, A. Bertuch, and U. Hansen. 1987. Specific stimulation of simian virus 40 late transcription in vitro by a cellular factor binding the simian virus 40 21-base-pair repeat promoter element. *Proc. Natl. Acad. Sci. USA* **84**:6025–6029.
31. Kioussis, D., E. F. Vanin, T. deLange, R. A. Flavell, and F. Grosveld. 1983. β -Globin inactivation by DNA translocation in $\gamma\beta$ -thalassemia. *Nature* **306**:662–664.
32. Kozak, M. 1987. An analysis of 5' non-coding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**:8125–8145.
33. Kozak, M. 1990. Downstream secondary structure facilitates recognition of initiation codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **87**:8301–8305.
34. Lemaire, P., C. Vesque, H. Schmitt, H. Stunnenberg, R. Frank, and P. Charnay. 1990. The serum-inducible mouse gene *Krox-24* encodes a sequence-specific transcriptional activator. *Mol. Cell. Biol.* **10**:3456–3467.
35. Lim, L. C., S. L. Swendeman, and M. Sheffery. 1992. Molecular cloning of the α -globin transcription factor CP2. *Mol. Cell. Biol.* **12**:828–835.
36. Lock, P., S. Ralph, E. Stanley, I. Boulet, R. Ramsay, and A. R. Dunn. 1991. Two isoforms of murine *hck* generated by utilization of alternative translational initiation codons exhibit different patterns of subcellular localization. *Mol. Cell. Biol.* **11**:4363–4370.
37. Lozzio, C. B., and B. B. Lozzio. 1975. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **45**:321–334.
38. Mehdi, H., E. Ono, and K. C. Gupta. 1990. Initiation of translation at CUG, GUG, and ACG codons in mammalian cells. *Gene* **91**:173–178.
39. Mellentin, J. D., S. D. Smith, and M. L. Cleary. 1989. Lyl-1, a novel gene altered by chromosomal translocation in T-cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. *Cell* **58**:77–83.
40. Moritz, T., P. Dutt, X. Xiao, D. Carstanjen, T. Vik, H. Hanenberg, and D. A. Williams. 1996. Fibronectin improves transduction of reconstituting hematopoietic stem cells by retroviral vectors: evidence of direct viral binding to chymotryptic carboxy-terminal fragments. *Blood* **88**:855–862.
- 40a. Morley, B. J., C. A. Abbott, and W. G. Wood. 1991. Regulation of human fetal and adult globin genes in mouse erythroleukemia cells. *Blood* **78**:1355–1363.
41. Nagpal, S., A. Zelent, and P. Chambon. 1992. RAR-beta 4, a retinoic acid receptor isoform is generated from RAR-beta 2 by alternative splicing and usage of a CUG initiator codon. *Proc. Natl. Acad. Sci. USA* **89**:2718–2722.
42. Nuez, B., D. Michalovich, A. Bygrave, R. Ploemacher, and F. Grosveld. 1995. Defective hematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375**:316–318.
43. Orkin, S. H. 1995. Regulation of globin gene expression in erythroid cells. *Eur. J. Biochem.* **231**:271–281.
44. Packham, G., M. Brimmell, and J. L. Cleveland. 1997. Mammalian cells express two differently localised Bag-1 isoforms generated by alternative translation initiation. *Biochem. J.* **328**:807–813.
45. Paszty, C., C. M. Brion, E. Mancini, H. E. Witkowska, M. E. Stevens, N. Mohandas, and E. M. Rubin. 1997. Transgenic knockout mice with exclusively human sickle hemoglobin and sickle cell disease. *Science* **278**:876–878.
46. Peabody, D. S. 1989. Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.* **264**:5031–5035.
47. Perkins, A. C., K. M. L. Gaensler, and S. H. Orkin. 1996. Silencing of human fetal globin expression is impaired in the absence of the adult β -globin gene activator protein, EKLF. *Proc. Natl. Acad. Sci. USA* **93**:12267–12271.
48. Perkins, A. C., A. H. Sharpe, and S. H. Orkin. 1995. Lethal β -thalassemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* **375**:318–322.
49. Poncz, M. P., P. Henthorn, C. Stoekert, and S. Surrey. 1989. Globin gene expression in hereditary persistence of fetal hemoglobin and $\delta\beta^0$ -thalassemia, p. 163–203. *In* N. McLean (ed.), *Oxford surveys of eukaryotic genes*. Oxford University Press, Oxford, United Kingdom.
50. Prats, H., M. Kaghad, A. C. Prats, M. Klagsburn, J. M. Lelias, P. Liauzun, P. Chalou, J. P. Tauber, F. Amalric, J. A. Smith, and D. Caput. 1989. High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. *Proc. Natl. Acad. Sci. USA* **86**:1836–1840.
51. Rowley, P. T., B. M. Ohlsson-Wilhelm, L. Wisniewski, C. B. Lozzio, and B. B. Lozzio. 1984. K562 human leukemia cell passages differ in embryonic globin gene expression. *Leuk. Res.* **8**:45–54.
52. Ryan, T. M., D. J. Ciavatta, and T. M. Townes. 1997. Knockout-transgenic mouse model of sickle cell disease. *Science* **278**:873–876.
53. Sabatino, D. E., A. P. Cline, P. G. Gallagher, L. J. Garrett, G. Stamatoyannopoulos, B. G. Forget, and D. M. Bodine. 1998. Substitution of the human β -spectrin promoter for the human γ -globin promoter prevents silencing of a linked human β -globin gene in transgenic mice. *Mol. Cell. Biol.* **18**:6634–6640.
54. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
55. Saris, C. J. M., J. Domen, and A. Berns. 1991. The pim-1 oncogene encodes two related protein-serine/threonine kinases by alternative initiation at AUG and CUG. *EMBO J.* **10**:655–664.
56. Shirra, M. K., Q. Zhu, H.-C. Huang, D. Pallas, and U. Hansen. 1994. One exon of the human *LSF* gene includes conserved regions involved in novel DNA-binding and dimerization motifs. *Mol. Cell. Biol.* **14**:5076–5087.
57. Stamatoyannopoulos, G., and A. W. Nienhuis. 1994. Hemoglobin switching, p. 107–156. *In* G. Stamatoyannopoulos, A. W. Nienhuis, P. J. Majerus, and H. Varmus (ed.), *The molecular basis of blood diseases*, 2nd ed. W. B. Saunders, Philadelphia, Pa.
58. Starck, J., R. Sarkar, M. Romana, A. Bhargava, A. L. Scarpa, M. Tanaka, J. W. Chamberlain, S. M. Weissman, and B. G. Forget. 1994. Developmental regulation of human γ - and β -globin genes in the absence of the locus control region. *Blood* **84**:1656–1665.
59. Sui, X., S. B. Krantz, M. You, and Z. Zhao. 1998. Synergistic activation of MAP kinase (ERK1/2) by erythropoietin and stem cell factor is essential for expanded erythropoiesis. *Blood* **92**:1142–1149.
60. Tagle, D. A., B. F. Koop, M. Goodman, J. L. Slightom, D. L. Hess, and R. T. Jones. 1988. Embryonic ϵ and γ globin genes of a prosimian primate (*Galago crassicaudatus*). *J. Mol. Biol.* **203**:439–455.
61. Taramelli, R., D. Kioussis, E. F. Vanin, K. Bartram, J. Groffen, J. Hurst, and F. G. Grosveld. 1986. $\gamma\delta\beta$ -Thalassemias 1 and 2 are the result of a 100 kbp deletion in the human β -globin cluster. *Nucleic Acids Res.* **14**:7017–7029.
62. Tuan, D., W. Solomon, Q. Li, and I. M. London. 1985. The “ β -like-globin” gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. USA* **82**:6384–6388.
63. Uv, A. E., C. R. L. Thompson, and S. J. Bray. 1994. The *Drosophila* tissue-specific factor grainyhead contains novel DNA-binding and dimerization domains that are conserved in the human protein CP2. *Mol. Cell. Biol.* **14**:4020–4031.
64. van der Ploeg, L. H., A. Konings, M. Oort, D. Roos, L. Bernini, and R. A. Flavell. 1980. $\gamma\beta$ -Thalassemia studies showing that deletion of the γ and δ genes influences β -globin expression in man. *Nature* **283**:637–642.
65. Vanin, E. F., P. S. Henthorn, D. Kioussis, F. Grosveld, and O. Smithies. 1983. Unexpected relationships between four large deletions in the human β -globin gene cluster. *Cell* **35**:701–709.
66. Wijgerde, M., J. Gribnau, T. Trimborn, B. Nuez, S. Philipsen, F. Grosveld, and P. Fraser. 1996. The role of EKLF in human β -globin gene competition. *Genes Dev.* **10**:2894–2902.
67. Wijgerde, M., F. Grosveld, and P. Fraser. 1995. Transcription complex stability and chromatin dynamics in vivo. *Nature* **377**:209–213.
68. Xiao, J. H., I. Davidson, H. Matthes, J. Garnier, and P. Chambon. 1991. Cloning, expression and transcriptional properties of the human enhancer factor TEF-1. *Cell* **65**:551–568.
69. Yang, Z., and J. D. Engel. 1994. Biochemical characterisation of the developmental stage- and tissue-specific erythroid transcription factor, NF-E4. *J. Biol. Chem.* **269**:10079–10087.
70. Yoon, J.-B., G. Li, and R. G. Roeder. 1994. Characterization of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro. *Mol. Cell. Biol.* **14**:1776–1785.