MOLECULAR AND CELLULAR BIOLOGY, Mar. 1994, p. 2201–2212
0270-7306/94/$04.00+0

Vol. 14, No. 3

Human GATA-3 *trans*-Activation, DNA-Binding, and Nuclear Localization Activities Are Organized into Distinct Structural Domains

ZHUOYING YANG,[1] LIN GU,[1] PAUL-HENRI ROMEO,[2] DOMINIQUE BORIES,[2] HOZUMI MOTOHASHI,[3] MASAYUKI YAMAMOTO,[3] AND JAMES DOUGLAS ENGEL[1]*

Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208-3500[1]; Department of Molecular Genetics and Hematology, INSERM U91, Hôpital Henri Mondor, 94010 Creteil, France[2]; and Department of Biochemistry, Tohoku University School of Medicine, Sendai 980, Japan[3]

Received 15 October 1993/Returned for modification 19 November 1993/Accepted 3 December 1993

GATA-3 is a zinc finger transcription factor which is expressed in a highly restricted and strongly conserved tissue distribution pattern in vertebrate organisms, specifically, in a subset of hematopoietic cells, in cells within the central and peripheral nervous systems, in the kidney, and in placental trophoblasts. Tissue-specific cellular genes regulated by GATA-3 have been identified in T lymphocytes and the placenta, while GATA-3-regulated genes in the nervous system and kidney have not yet been defined. We prepared monoclonal antibodies with which we could dissect the biochemical and functional properties of human GATA-3. The results of these experiments show some anticipated phenotypes, for example, the definition of discrete domains required for specific DNA-binding site recognition (amino acids 303 to 348) and *trans* activation (amino acids 30 to 74). The signaling sequence for nuclear localization of human GATA-3 is a property conferred by sequences within and surrounding the amino finger (amino acids 249 to 311) of the protein, thereby assigning a function to this domain and thus explaining the curious observation that this zinc finger is dispensable for DNA binding by the GATA family of transcription factors.

Human GATA-3 (hGATA-3) is a 444-amino-acid vertebrate transcription factor protein which binds with high affinity to related consensus DNA sites (GATA, GATTA, and GATCT) through a two-C_4-zinc-finger DNA-binding domain (23, 24, 31, 52, 54). GATA-1, the founding member of this multigene family, was originally detected in and cloned from erythroid lineage cells (1, 7, 8, 30, 36, 37, 49). Cloning of the related factors, GATA-2 and GATA-3, was reported shortly thereafter (52), and newer family members have since been identified (2, 22).

The basis for classification of these proteins arises from the highly conserved tissue-restricted distribution patterns of individual GATA family members in different vertebrate organisms. The amino acid sequence of the DNA-binding domain is over 85% identical among different GATA family members and is also highly conserved among vertebrate species (35), while regions outside the DNA-binding domains vary from extremely high similarity between species (e.g., 92% for GATA-3 [24]) to low identity (e.g., about 40% for chicken GATA-1 [cGATA-1], mouse GATA-1 [mGATA-1], and hGATA-1 [7, 47, 49, 57]).

GATA-3 was first shown to be abundantly expressed in T lymphocytes and in the embryonic brain (16, 20, 24, 52). A number of target genes for GATA-3-directed transcriptional activity have been identified in T lymphocytes, which only express GATA-3 among the members of this family (24, 26, 52). Both the human and murine T-cell receptor (TCR) δ and α gene enhancers were shown to bind GATA-3, and mutational analysis showed that the GATA-binding sites were required for tissue-specific TCR δ enhancer activity (16, 20, 24, 38, 39). hGATA-3 also binds to multiple sites within the human immunodeficiency virus type 1 long terminal repeat and is required for maximal stimulation of human immunodeficiency virus long terminal repeat-directed transcription in T cells (54). GATA-3 is abundantly expressed during T-lymphocyte differentiation both in mature $CD4^+$ $CD8^-$ and $CD4^-$ $CD8^+$ and in less mature $CD4^+$ $CD8^+$ T cells (26). Given that GATA-binding sites have been identified in the regulatory regions of a growing number of T-cell-specific genes, we infer that GATA-3 may be one of the transcription factors responsible for T-cell identity and/or determination.

More recently, GATA-2 and GATA-3 have also been shown to be abundantly expressed in placental trophoblasts, in which the human gonadotropin α-subunit promoter (30a) and the mouse placental lactogen I gene promoter (27a) are regulated by a GATA factor. GATA-3 is expressed at the earliest stages of *Xenopus*, chicken, and murine embryonic development (24, 26, 56) and in the developing vertebrate central nervous system (25), peripheral nervous system (11b), and kidney (25a). Homozygous mutant mice lacking mGATA-3 fail to survive gestation, consistent with the expectation that GATA-3 fulfills a vital regulatory function during vertebrate embryogenesis (27).

Structural and functional studies of GATA-1 have shown that the amino-terminal segment of the polypeptide contributes to the transcriptional activation function of the protein, while two related zinc fingers, lying somewhat closer to the carboxyl terminus of the protein, determine the DNA-binding affinity and specificity (29, 53). The two fingers of GATA-1 can be functionally subdivided; the amino finger (which we refer to here as finger I) contributes to the specificity and stability of DNA binding, while the carboxyl finger (finger II) is absolutely required for recognition of the GATA consensus motif (29, 34, 53). Since the amino acid sequence of the entire family of factors is conserved throughout the finger domain, we would

* Corresponding author. Mailing address: Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, IL 60208-3500. Phone: (708) 491-5139. Fax: (708) 467-2152. Electronic mail address: d-engel@nwu.edu.

anticipate similar results for hGATA-3 DNA-binding specificity and affinity.

To elicit their functions, transcription factors must encode nuclear targeting signals to allow the protein to traverse the nuclear membrane. The essence of the simplest nuclear localization sequence appears to be a short string of basic amino acids (4), and the control of nuclear localization has been shown to play a critical role in the activity of several transcription factors. The GATA proteins have not been analyzed with regard to whether or not nuclear localization plays a functional role in the transcriptional regulation elicited by these proteins.

Since previous studies have usually assayed the biochemical activity of different members of the GATA factor family by indirect DNA binding assays (electrophoretic gel mobility shift assays [EGMSA]), we wished to examine the biological properties of GATA-3 by using a direct assay for the activities and/or cellular distribution of the protein. To address the relationship between the primary sequence of hGATA-3 and its functional properties, we first prepared monoclonal antibodies (MAbs) recognizing the bacterially expressed protein. We then undertook a systematic mutational analysis of hGATA-3, examining in-frame deletions within the hGATA-3 coding region by expressing each of the mutated proteins in eucaryotic cells. The effects of the mutations were assessed with the anti-hGATA-3 MAbs by separately examining the ability of various mutated hGATA-3 proteins to bind to DNA, to activate transcription, and to direct localization of the factor to the nucleus. We find that the *trans*-activating domain is localized within the amino-terminal portion of the protein while the DNA-binding domain resides within the carboxyl finger. Using immunolocalization assays, we show that the hGATA-3 protein is found within the nuclei, both in cells in which it is normally expressed and in transfected cells. Finally, we showed that amino acid (aa) residues 249 to 311 are required for nuclear localization of the hGATA-3 protein.

MATERIALS AND METHODS

Cell lines. Human T-lymphocyte cell lines Jurkat and HuT78 and murine T-lymphocyte cell line BW5147.3 were purchased from the American Type Culture Collection. YN79 (a subline of the Y79 human retinoblastoma line), NB-1 (neuroblastoma), and HEL (erythroleukemia) lines were generously provided by S. Tsuchiya. MEL (murine erythroleukemia) and C1300 (neuroblastoma) cell lines were provided by Vikram Patel and Bernard Mirkin, respectively. Transformed quail fibroblast (QT6) cells, which do not express GATA factors, were used as previously described (24, 52).

Generation of MAbs recognizing GATA-3. The first 785 nucleotides encoding the amino-terminal portion of the hGATA-3 cDNA, ending at aa 264 (the first cysteine encountered in the amino finger [20, 24]), were amplified by PCR using oligonucleotide primers containing synthetic restriction sites for convenient *Bam*HI-*Eco*RI insertion into the bacterial expression plasmid pGEX-A (a gift of Jurgen Kun, Institut für Genetik der Universität Köln). The PCR product was verified by DNA sequencing. Cells containing the bacterial expression plasmid were grown to mid-log phase, and the glutathione *S*-transferase (GST)/hGATA-3 fusion peptide was induced by the addition of isopropylthiogalactopyranoside (IPTG) to 0.5 mM for 3 h at 37°C. Cells were then lysed, and the fusion protein was bound to and recovered from a glutathione column as previously described (46).

Approximately 100 μg of GST/hGATA-3 was injected intraperitoneally into each mouse, using Hunter's Titermax adjuvant (CytRx Corp., Norcross, Ga.). After three consecutive injections with 100 μg of fusion protein every 3 weeks, animals were assayed by enzyme-linked immunosorbent assay (ELISA) for the ability of the serum to recognize GST/hGATA-3. Four weeks after the final injection, spleen cells were fused to NSO-1 myeloma cells as described previously (10). Positive hybridoma wells were initially identified by ELISA using the fusion protein as the fixed antigen and subsequently by biochemical assays using native hGATA-3 protein (see below).

hGATA-3 mutations. All deletions were prepared either by using convenient restriction enzyme sites to excise specific fragments or by PCR, as shown in Table 1, starting with the base plasmid Rous sarcoma virus (RSV)/hGATA-3 described previously (24). In either case, mutations were chosen to result in in-frame deletions, and the integrity of each mutation was confirmed by DNA sequencing. All of the deletion mutations examined are described in Table 1 and are numbered consecutively in the order in which they are described here.

To prepare the GAL4/hGATA-3 fusion proteins, either PCR products or restriction fragments of the hGATA-3 coding region were subcloned in frame into the multiple cloning sites 3′ to the GAL4 DNA-binding and dimerization domain (aa 1 to 147) of plasmid pSG424 (28). These mutations were also confirmed to be in frame by DNA sequencing. The hGATA-3 mutations examined as GAL4 fusion chimeras are described in Table 2.

EGMSA and supershift assays. For EGMSA (9, 11) of overexpressed hGATA-3 protein, whole-cell extracts were prepared from hGATA-3-transfected QT6 cells disrupted with high-salt lysis buffer (500 mM NaCl, 1% Triton X-100, 50 mM Tris-HCl [pH 7.5], phenylmethylsulfonyl fluoride [PMSF]) and partially purified by double-stranded calf thymus DNA-cellulose chromatography (10). The extracts were then incubated with radiolabeled oligonucleotide TδE4, which corresponds to the GATA sequences within the human TCR δ gene enhancer footprint 4 (24, 38). Gel shift assays were performed as described previously in a final buffer consisting of 80 mM NaCl, 25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES; pH 7.7), 10 mM dithiothreitol [DTT], 5 mM $MgCl_2$, 100 mg of bovine serum albumin per ml, and 2.5% Ficoll (52). For the supershift assay, 2 μl of supernatant from each ELISA-positive hybridoma was added to the gel shift reaction 20 min after initiation of the GATA-3/TδE4 DNA binding reaction at 4°C, and then the assay was performed as described above.

The mouse thymus cell, Jurkat cell, and NB-1 cell nuclear extracts were prepared first by addition of lysis buffer (20 mM HEPES [pH 7.7], 20% glycerol, 10 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% Triton X-100, 1 mM DTT, 1 mM PMSF, leupeptin, and pepstatin). The nuclear pellets were then treated with nuclear extraction buffer (lysis buffer plus 400 mM NaCl). The nuclear extracts were then dialyzed against 20 mM HEPES (pH 7.9)–20% glycerol–100 mM KCl–0.2 mM EDTA–1 mM DTT–1 mM PMSF. The probe used was a GATA site found within the vasoactive intestinal peptide (VIP) gene promoter (5′-GATCCAAAATGTA**AGATAA**GA GGAAATTT-3′); the mutated VIP GATA site substituted the sequence 5′-**ACCTAA**-3′ within the GATA consensus nucleotides shown in boldface in the wild-type sequence. For EGMSA using the GAL4/hGATA-3 fusion protein, a probe containing five copies of the GAL4-binding site excised from pSG5CAT (28) by *Hin*dIII and *Kpn*I was radiolabeled by using polynucleotide kinase.

The protein extracts prepared here are somewhat unstable over time even when maintained at −70°C, and therefore the gel shift and immunoprecipitation experiments were usually carried out immediately after lysis. Nonetheless, slight exper-

TABLE 1. Strategy for generating hGATA-3 deletion constructs[a]

hGATA-3 deletion	Method	Start point (primer or enzyme)	End point (primer or enzyme)
2 (d2–31)	PCR/*Nco*I	5′-aattc(203)CATGGAG(284)C TCAGCCACTCCTACATG	5′-ctaacc(1532)CATGGCGG TGAC(1520)
3 (d2–59)	PCR/*Nco*I	5′-aattc(203)CATGGAG(378)C ACGTCCCGCCCTAC	5′-ctaacc(1532)CATGGCGG TGAC(1520)
4 (d73S–133)		*Kpn*I (T4) at 428	*Acc*I (K) at 598
5 (d132–214)		*Acc*I (K) at 598	*Acc*I (K) at 843
6 (d58Q–82)	*Kpn*I/*Bal* 31[b]	Delete to nt 380	Delete to nt 446
7 (d29–128)		*Sma*I at 289	*Sma*I at 583
8 (d29A–228)		*Ava*I (K) at 287	*Ava*I (K) at 882
9 (d173–214)		*Nae*I at 722	*Sal*I (K) 842
10 (d214–241)		*Hinc*II at 844	*Bgl*I (T4) at 922
11 (d348–395)		*Hin*dIII (K) at 1245	*Eag*I (K) at 1384
12 (d395–443)	PCR/*Nco*I	5′-gac(203)CATGGAGGTGACt GCGGACCA(213)	5′-tgatc(1532)CAT(1388)GGCC GGGTTAAACGAGCT
13 (d249–303)	N terminal	5′-gac(203)CATGGAGGTGACt GCGGACCA(213)	5′-(950)GCATCCGAAGCCGGTGGG(933)
	C terminal	5′-(1110)AAGCCCAAGCGAAGGCTG(1127)	5′-ctaacc(1532)CATGGCGG TGAC(1520)
14 (d269–303)	N terminal	5′-gac(203)CATGGAGGTGACt GCGGACCA(213)	5′-(1011)TGCCCGACAGTTCACACA(993)
	C terminal	5′-(1110)AAGCCCAAGCGAAGGCTG(1127)	5′-ctaacc(1532)CATGGCGG TGAC(1520)
15 (d309–329)		*Pst*I (T4) at 1136	*Pfi*mI (T4) at 1189
16 (d343–355)	*Hin*dIII/*Bal* 31[b]	Delete to nt 1232	Delete to nt 1266
17 (d269–343)	N terminal	5′-gac(203)CATGGAGGTGACt GCGGACCA(213)	5′-(1011)TGCCCGACAGTTCACACA(993)
	C terminal	5′-(1230)GGGCTCTACTACAAGCTT(1248)	5′-ctaacc(1532)CATGGCGG TGAC(1520)
18 (214–347)		*Sal*I (K) at 842	*Hin*dIII (K) at 1245
19 (249–258)	N terminal	5′-gac(203)CATGGAGGTGACt GCGGACCA(213)	5′-(950)GCATCCGAAGCCGGTGGG(933)
	C terminal	5′-(975)AGCACAGAAGGCAGGGAG(992)	5′-ctaacc(1532)CATGGCGG TGAC(1520)
20 (249–311)	N terminal	5′-gac(203)CATGGAGGTGACt GCGGACCA(213)	5′-(950)GCATCCGAAGCCGGTGGG(933)
	C terminal	*Pst*I (T4) at 1136	*Nco*I at 1528
21 (d249–258 + 303–311)	N terminal, 19 as template	5′-gac(203)CATGGAGGTGACt GCGGACCA(213)	5′-(1113)CTTAATGAGGGGCCGGTT(1095)
	C terminal	*Pst*I (T4) at 1136	*Nco*I at 1528

[a] The deletion mutations are numbered consecutively in the order in which they were used; designations in parentheses correspond to each deletion. T4, T4 polymerase followed by restriction enzyme sites to generate blunt ends when necessary; K, Klenow enzyme. For primers used in PCRs, the numbers are nucleotide numbers as described by Ko et al. (24). Capital letters represent nucleotides remaining in the final constructs.

[b] The RSV/hGATA-3 construct was first digested with *Kpn*I or *Hin*dIII, then treated with *Bal* 31, and sequenced for the in-frame deletions. nt, nucleotide.

imental variation due to proteolysis might be reflected in the subtle differences observed between different samples in the EGMSA and immunoprecipitation studies.

Immunoprecipitation. After logarithmic growth for several generations, cells were replated in Cys- and Met-deficient Dulbecco's modified Eagle medium for 30 min, labeled with Trans [^{35}S]Met/Cys (ICN) for 3 h, and then lysed with high-salt lysis buffer (see above). Transfected QT6 cells were labeled in the same way 2 days after transfection. The extracts were then incubated with supernatant MAb and subsequently with protein A/G beads (Schleicher & Schuell). The beads were then sequentially washed with high-salt lysis buffer and sodium dodecyl sulfate (SDS) wash buffer (0.1% Triton X-100, 0.02% SDS, 150 mM NaCl, 50 mM Tris-HCl [pH 7.5], 5 mM EDTA, PMSF) and finally resuspended in SDS loading buffer (50 mM Tris-HCl [pH 6.8], 2% SDS, 40 mM DTT, 10% glycerol, 0.1% bromphenol blue). The samples were separated by electrophoresis on an SDS–10% polyacrylamide gel after boiling. The gel was treated afterward by incubation with 100% acetic acid, fixed in 22% 2,5-diphenyloxazole (PPO) in acetic acid and finally extensively washed in water. The gel was then dried and exposed to X-ray film overnight.

***trans*-activation assays.** QT6 cells were grown in Dulbecco's modified Eagle medium containing 2% tryptose phosphate broth, 5% newborn calf serum, and 2% chicken serum and were transfected by the calcium phosphate procedure (13) with 2 μg of the RSV/hGATA-3 wild type (24) or deletion mutations (as indicated in the figure legends) together with 8 μg of Eδ3GH, which has three copies of TδE4 binding sites directing transcription of the growth hormone (GH) gene through a minimal promoter (24). Forty percent of the cells from a confluent plate were used for each transfection. At 16 h after transfection, the cells were rinsed with phosphate-buffered saline (PBS) and resuspended in growth medium. Supernatants were assayed the following day for secreted human GH, using the Allegro hGH kit (Nicholas Institute Diagnostics, San Juan Capistrano, Calif.). *trans*-activation activity was calculated as described previously (52).

Previous experience indicated that the results of cotransfection assays using GH as reporter gene are more consistent from experiment to experiment than the results of experiments using either chloramphenicol acetyltransferase (CAT) or luciferase (LUC) enzymatic activity. This may be because in the GH assay, one simply samples the growth medium, thereby

TABLE 2. Strategy for generating GAL4/hGATA-3 fusion constructs[a]

Construct	Cloning site in pSG424	Start point of hGATA-3	End point of hGATA-3
2	*Sma*I	*Nco*I (K) at 202	*Sma*I at 289
3	*Sma*I	*Nco*I (K) at 202	*Kpn*I (T4) at 428
4	*Sma*I	*Nco*I (K) at 202	*Sal*I (K) at 842
5[b]	*Sma*I	5′-gac(203)CATGGAGGTGACtG CGGACCA(213)	5′(933)TGTGTGAACTGTGGGGCA
6	*Sma*I	*Nco*I (K) at 202	*Hin*dIII (K) at 1248
7	*Sma*I	*Nco*I (K) at 202	*Nco*I (K) at 1528
8[c]	*Sma*I	*Nco*I (K) at 202	*Nco*I (K) at 1528
9	*Sma*I	*Sma*I at 289	*Kpn*I (T4) at 428
10	*Sma*I	*Sma*I at 289	*Bam*HI (K) at 477
11	*Sma*I	*Sma*I at 289	*Sma*I at 583
12	*Bam*HI (K)	*Kpn*I (T4) at 428	*Sal*I (K) at 842
13	*Sma*I	*Acc*I (K) at 598	*Acc*I (K) at 843

[a] hGATA-3 fragments were cloned into the restriction enzyme sites in pSG424 as indicated. Abbreviations are as for Table 1.
[b] The hGATA-3 fragment was generated by PCR.
[c] The parental plasmid for the hGATA-3 portion is construct 18 of hGATA-3 from Fig. 4A.

avoiding cell lysis and thus minimizing potential complications from differential protein extraction or degradation. On the other hand, it is relatively difficult to internally control for GH transfection efficiency, since one may encounter significant variability in the activity of secreted (GH) and endogenous (CAT or LUC) proteins when analyzing different transfection samples. We therefore repeated each GH reporter transfection experiment on at least three occasions for the data reported here.

In other experiments, a CAT instead of GH reporter gene was used. At 48 h after transfection, cell lysates were prepared by freeze-thaw (40). A LUC assay (internal transfection efficiency control) was carried out first, and CAT assays were performed with aliquots of extract containing equal quantities of LUC activity. The CAT assay results were quantified by determining the amount of acetylated [^{14}C]chloramphenicol produced in the enzymatic assay (12); the degree of conversion was quantitated on a Molecular Dynamics PhosphorImager.

Subcellular immunolocalization. For immunohistochemical staining, human HuT78, NB-1, YN79, or HEL cells were displayed on glass slides by using a Cytospin III (Shandon), fixed in 1% paraformaldehyde for 5 min, and then fixed in cold acetone for 5 min. After incubation in 2% sheep serum, specimens were reacted with MAb 31 or 35 overnight at 4°C. After being washed in PBS, the samples were incubated with horseradish peroxidase-conjugated F(ab′)$_2$ fragment of anti-mouse immunoglobulin G (1:100 dilution; Amersham) overnight at 4°C. Diaminobenzidine was used as the chromogen. Nuclei were counterstained with methyl green for 1 h.

For immunofluorescence staining, QT6 cells transfected with hGATA-3 or hGATA-3 deletion constructs were treated with trypsin and replated onto coverslips 24 h prior to staining. The coverslips were fixed in 1% paraformaldehyde, permeabilized in −20°C acetone, and then incubated with anti-hGATA-3 MAb or a control MAb (which binds to GST) for 1 h. The coverslips were then washed with PBS before addition of secondary fluorescein isothiocyanate-labeled mouse immunoglobulin G (Jackson Laboratory). Cell staining was visualized and photographed on a Bio-Rad MRC 600 confocal fluorescence microscope.

RESULTS

MAbs specifically recognize GATA-3. To independently assess the biochemical properties of hGATA-3, we generated MAbs recognizing the transcription factor by injecting mice with bacterially expressed, affinity-purified GST/hGATA-3 fusion protein. This fusion protein contained the amino-terminal portion of the hGATA-3 protein, ending (at the carboxyl terminus) at the first cysteine residue defining finger I of the protein (24) (Materials and Methods).

Hybridoma clones which initially tested positive in ELISA were subsequently examined by supershift EGMSA, in which partially purified hGATA-3 protein recovered from transfected QT6 quail fibroblast cells (which express no endogenous GATA proteins [24, 52]) was mixed with radiolabeled oligonucleotide containing the TδE4 binding site (24) as well as supernatant antibody taken from each ELISA-positive hybridoma culture (Materials and Methods). Four positive hybridoma wells were identified, and two of these (clones 31 and 35) were subcloned to generate monoclonal lines.

MAb 31 was tested in immunoprecipitation experiments with extracts prepared from a human T-cell line, Jurkat, which expresses abundant hGATA-3 mRNA and protein (54) to test whether it could recognize native hGATA-3 protein. MAb 31 indeed specifically reacted with a protein of approximately 50 kDa, in good agreement with the predicted size of the hGATA-3 polypeptide (24), whereas a control MAb failed to recognize any similar protein (Fig. 1A, lanes 1 and 2). The epitope recognized by MAb 31 was later found to lie within the first 30 aa of hGATA-3.

To determine the specificity of MAb 31, we performed similar immunoprecipitation experiments using extracts prepared from BW5147.3 (a mouse T-cell line) and C1300 cells (a mouse neuroblastoma cell line), both of which have been shown to abundantly express murine GATA-3 mRNA (11a, 24). The results (Fig. 1A, lanes 4 and 7) show that MAb 31 recognizes a 50-kDa protein in both murine cell types in comparison with the control antibody (Fig. 1A, lanes 3 and 5). In contrast, MAb 31 fails to recognize mGATA-1 (from murine erythroleukemia cells; Fig. 1A, lanes 8 and 9) or mGATA-2 (which is also abundantly expressed in C1300 neuroblastoma cells; Fig. 1A, lane 6). Supershift experiments confirmed that MAb 31 specifically recognizes the protein which binds to a GATA-containing oligonucleotide in extracts prepared from mouse thymus cells, human T-lymphocyte Jurkat cells, and human neuroblastoma NB-1 cells (Fig. 1B, lanes 2, 6, and 10). When bacterially expressed GST/cGATA-1, -2, and -3 proteins were examined in a supershift EGMSA experiment, both MAb 31 and MAb 35 reacted only with

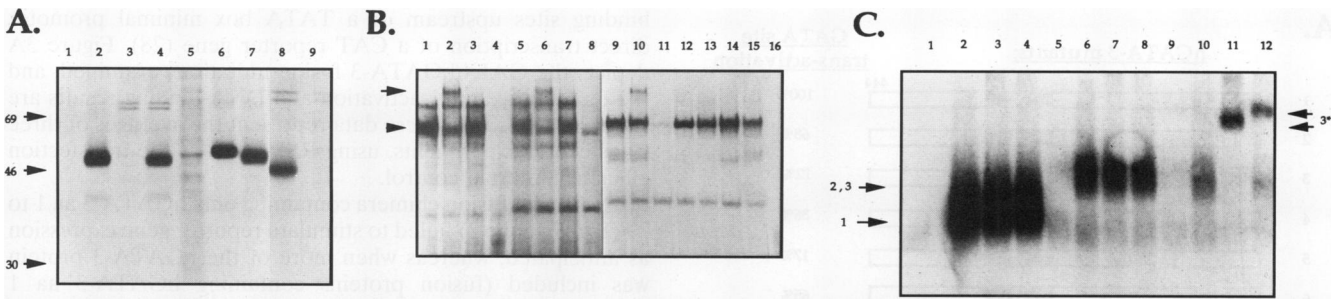

FIG. 1. Anti-hGATA-3 MAbs specifically recognize human, mouse, and chicken GATA-3 protein. (A) Immunoprecipitation of cell extracts. A MAb recognizing GST (lanes 1, 3, and 5), murine GATA-2 (14a) (lane 6), or murine GATA-1 (19) (lane 8) or anti-hGATA-3 MAb 31 (lanes 2, 4, 7, and 9) was used in immunoprecipitation reactions. The cells examined were from human (Jurkat; lanes 1 and 2) or mouse (BW5147.3, lanes 3 and 4) T-lymphocyte cell lines, the C1300 mouse neuroblastoma cell line (lanes 5 to 7), or the MEL (mouse erythroleukemia; lanes 8 and 9) cell line. Arrows on the left depict the gel mobility positions of the protein size markers. (B) EGMSA supershift of T-cell and neuroblastoma cell extracts. Extracts were prepared from mouse thymus cells (lanes 1 to 4), human T-lymphocyte Jurkat cells (lanes 5 to 8), and human neuroblastoma NB-1 cells (lane 9 to 16) as described in Materials and Methods. The binding reactions were carried out with a radiolabeled GATA-binding site from the VIP gene promoter (Materials and Methods) as a probe together with no antibody (lanes 1, 5, and 9), MAb 31 (lanes 2, 6, and 10), negative control mouse serum (lanes 3, 7, and 15), double-stranded, unlabeled GATA-binding site as a specific competitor (lanes 4, 8, and 11), or sense-strand (lane 12), antisense-strand (lane 13), or mutated (lane 14) VIP-1 GATA site as a cold competitor. Lane 16 is probe only. The lower arrow points to the probe/mGATA-3 protein complex, and the upper arrow represents the ternary complex of the probe, mGATA-3 protein, and the antibody. (C) EGMSA supershift of chicken GST/GATA-1, -2, and -3 proteins. Bacterially expressed, affinity-purified GST/cGATA-1 (lanes 2 to 4), GST/cGATA-2 (lanes 6 to 8), or GST/cGATA-3 (lanes 10 to 12) fusion proteins (23) were used in conjunction with oligonucleotide MαP (49) in this supershift experiment. Supernatants from MAb 31 (lanes 3, 7, and 11), MAb 35 (lanes 4, 8, and 12), or growth medium alone (lanes 2, 6, and 10) were added to each of the EGMSA reactions. Lanes 1, 5, and 9 contain no added protein. The arrows and numbers on the left indicate the positions of GST/cGATA-1, -2, or -3 and MαP probe complex in the EGMSA assay, and the arrow labeled 3* on the right represents the ternary complex of DNA/GATA-3 protein which has reacted with the antibody.

cGATA-3 (Fig. 1B, lanes 11 and 12), not with cGATA-1 or cGATA-2 (Fig. 1B, lanes 2 to 4 and 6 to 8). Taken together, the EGMSA supershift and immunoprecipitation experiments show that both anti-hGATA-3 MAbs 31 and 35 recognize human, mouse, and chicken GATA-3 proteins but fail to bind to chicken or mouse GATA-1 or GATA-2. Given the high degree of sequence identity between the murine and human GATA-1 and GATA-2 proteins (47, 57), we anticipate that these antibodies will not recognize hGATA-1 or hGATA-2.

Deletion analysis of hGATA-3 reveals two regions required for *trans* activation. A series of deletion mutations within the coding region of the hGATA-3 protein was generated as described in Materials and Methods and cotransfected into QT6 cells along with the Eδ3GH reporter plasmid, containing three copies of the T-cell receptor δE4 GATA sites driving expression of a GH cDNA from a minimal (TATA box) promoter (24). After collection of the supernatants to assay secreted GH levels, cotransfected cells were then lysed to prepare extracts for assessing transfected hGATA-3 DNA binding by EGMSA. The δE4 oligonucleotide was used as a gel shift probe with the hGATA-3 protein expressed from each of the deletion mutants, and each was found to accumulate in the transfected cells and to bind to the δE4 GATA sites (Fig. 2B); the mobility differences reflect the sizes of each of the hGATA-3 deletion mutated proteins. Although several of the hGATA-3 deletion mutated proteins appear to accumulate less stably than others, generally each transfection results in abundant accumulation of GATA-3 protein. Thus, GH detected in the medium is a direct reflection of the *trans*-activation function of hGATA-3; these results are summarized in Fig. 2A (averages of four independent experiments), in which the level of GH produced by *trans* activation with wild-type hGATA-3 protein was arbitrarily assigned a level of 100%.

Removing 29 of the amino-terminal amino acids (Fig. 2A, line 2) caused only a modest decrease in the activation ability of hGATA-3, while deletion of an additional 28 aa (Fig. 2A, line 3) resulted in a more significant loss of activity (>5-fold), indicating that aa 31 to 59 (Act I) are critical for *trans* activation by the hGATA-3 protein. Removing aa 132 to 214 (Act II; Fig. 2A, line 5) also results in major diminution of activity. Deletion of aa 29 through 128 gives a result comparable to that of the protein missing Act I alone, indicating that the sequences between aa 60 and 128 do not significantly contribute (>3-fold) to *trans* activation, while deletion of aa 29 to 228 is equivalent to deletion of both presumptive *trans*-activation domains (Act I plus Act II). Independent deletions encompassing aa 58 to 82, 173 to 214, 214 to 241, 348 to 395, and 395 to 422 (Fig. 2A, lines 6, 9, 10, 11, and 12, respectively) do not result in significant (>3-fold) alterations in the *trans*-activation activity compared with wild-type hGATA-3. We thus tentatively concluded from these data that the *trans*-activation domains of the hGATA-3 protein were within the GATA-3 coding sequence between residues 29 and 228 and that two functional subdomains (Act I [aa 31 to 59] and Act II [aa 132 to 214]) acted either independently or as a single cooperative domain to confer hGATA-3 *trans* activation.

hGATA-3 aa 30 to 74 are sufficient to confer transcriptional activation to a heterologous DNA-binding domain. To test whether Act I and Act II each function independently in hGATA-3 *trans* activation, we generated an additional series of constructs which would produce GAL4/hGATA-3 fusion proteins to examine whether the presumptive *trans*-activation domains identified by deletional analysis were also competent in *trans* activation of a reporter gene when linked to a heterologous DNA-binding domain. Parts, or all, of the hGATA-3 coding sequence were cloned in frame (and 3′) to the GAL4-encoded sequences into plasmid pSG424, which contains the first 147 aa of the GAL4 DNA-binding and dimerization domain ($GAL4_{1-147}$ [21]) directed by the simian virus 40 early promoter (28). The reporter plasmid used in this series of experiments was pSG5CAT, in which five GAL4-

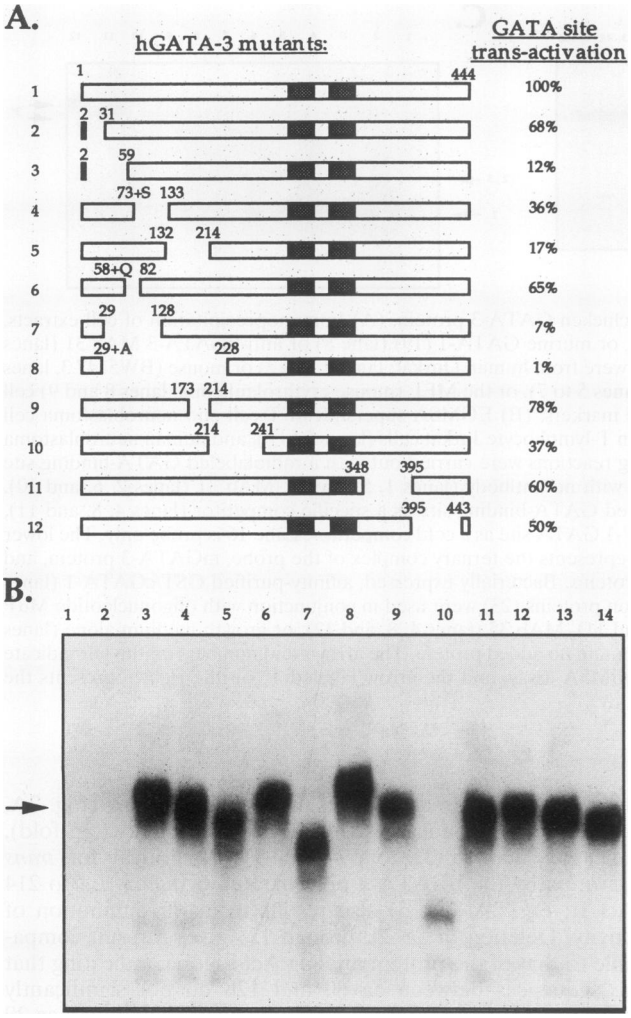

FIG. 2. Deletion mutations identify two domains required for *trans* activation by hGATA-3. (A) Schematic structure of hGATA-3 protein deletions and cotransfection *trans*-activation results. The top diagram represents the wild-type hGATA-3 protein, beginning at aa 1 (at the N terminus) and extending to aa 444 (at the C terminus) (20, 24). The two stippled regions represent the individual C_4 zinc fingers (24). Subsequent diagrams represent various hGATA-3 deletion mutations, whose construction details are described in Table 1 and are identified by numbers at the left. Numbers above each construct indicate either the first or the last amino acid of each deletion. On this and all subsequent diagrams, capital letters indicate extra amino acids generated during the various cloning steps which are not encoded in the wild-type hGATA-3 protein. On the right, the numbers represent the relative *trans*-activation ability (averages of four independent cotransfection experiments) of each of the mutated proteins in comparison with wild-type hGATA-3 (line 1). The calculated standard deviations of transfections for the constructs in percentages are as follows: 1, 15.1; 2, 16.1; 3, 3.0; 4, 7.8; 5, 14.7; 6, 19.7; 7, 3.1; 8, 0.6; 9, 7.6; 10, 8.5; 11, 17.8; and 12, 10.1. (B) DNA binding of the hGATA-3 mutated proteins. QT6 cells cotransfected with RSV/hGATA-3 or each of the constructs encoding the mutated proteins were used to prepare cellular extracts for EGMSA (Materials and Methods). The probe used in this case was radiolabeled TδE4 oligonucleotide, containing two GATA-binding sites (24). The hGATA-3/GATA-binding site complexes are indicated on the left (arrows). Lanes: 1, labeled oligonucleotide alone; 2, a reaction with extracts prepared from QT6 cells transfected with antisense hGATA-3; 3 to 14, gel shift results with extracts prepared from QT6 cells transfected with each of the RSV/hGATA-3 mutations corresponding to lines 1 to 12 of panel A. The different positions of the shifted bands reflect changes in the size and/or charge of the mutated hGATA-3 proteins.

binding sites upstream of a TATA box minimal promoter direct transcription of a CAT reporter gene (28). Figure 3A depicts the GAL4/hGATA-3 fusion mutations examined, and the quantitative *trans*-activation and DNA-binding results are shown on the right. These data represent the averages of three transfection experiments, using RSV-LUC as the transfection efficiency internal control.

The GAL4 fusion chimera containing only hGATA-3 aa 1 to 29 (Fig. 3A, line 2) failed to stimulate reporter gene expression as anticipated, whereas when more of the hGATA-3 protein was included (fusion proteins containing hGATA-3 aa 1 through 74, 1 through 214, and 1 through 269; Fig. 3A, lines 3 to 5), each could direct high-level *trans* activation. Surprisingly, when either full-length hGATA-3 or aa 1 through 348 were linked to $GAL4_{1-147}$, no *trans* activation of the reporter gene was detected (Fig. 3A, lines 6 and 7). Examination of a fusion protein which includes the full-length hGATA-3 with only the zinc finger domain removed (Fig. 3A, line 8) showed that reporter gene expression was substantially restored, indicating that the zinc finger region of hGATA-3 is responsible for loss of *trans*-activation activity in a GAL4 chimera which includes the hGATA-3 zinc finger region.

The blocking of hGATA-3 *trans* activation by its own zinc finger domain in the GAL4/hGATA-3 fusion protein chimera could be due to (i) a direct *trans*-activation suppressor function encoded within the hGATA-3 zinc finger domain, (ii) the presence of a second DNA-binding domain within the fusion protein which blocks specific DNA binding of $GAL4_{1-147}$ to the GAL4-binding sites in the reporter gene; or (iii) the possibility that fusion proteins containing two DNA-binding domains are intrinsically unstable and therefore fail to accumulate in cells. To distinguish among these various possibilities, we first performed indirect immunoprecipitation analysis of the proteins recovered from transfected cells to determine whether the mutated proteins were stably expressed; all of the chimeric proteins were found to abundantly accumulate in transfected cells (Fig. 3B).

To distinguish among the remaining possibilities, we then examined whether these transfected chimeric proteins were capable of binding to a GAL4 site in vitro (Fig. 3C); the probe in this case contained five repeats of the GAL4-binding site (excised from pSG5CAT). The GAL4 DNA-binding domain alone and each of the fusion proteins bind to DNA, except those (full-length hGATA-3 or aa 1 to 348) which included the intact zinc finger DNA-binding domain of hGATA-3 (Fig. 3C, lanes 1 to 7). The reciprocal lack of binding of fusion proteins containing both intact DNA-binding domains to a GATA consensus sequence is also true: none of the chimeras which contained both intact DNA-binding domains bind to a consensus GATA site (Fig. 3C, lanes 8 to 15). These data indicate that the failure of fusion chimeras containing both GATA-3 and GAL4 DNA-binding domains to *trans* activate a GAL4-dependent reporter gene is due, at least in part, to the inability of fusion proteins harboring two independent DNA-binding domains to bind to either GAL4- or GATA-binding sites.

On the basis of results of the 3′ deletion *trans*-activation studies (Fig. 3A, lines 1 to 8), smaller segments of the hGATA-3 cDNA sequence were examined in order to generate GAL4/hGATA-3 fusion proteins, each containing a presumably independent *trans*-activation domain of hGATA-3. As illustrated (Fig. 3A, lines 9 to 13), a fusion protein containing hGATA-3 aa 30 to 74 (including Act I) is capable of stimulating reporter gene activation, while the addition of aa 75 to 93 (line 10) or 75 to 127 (line 11) only modestly augments that activity. However, a fusion chimera containing hGATA-3 aa 133 to 214 (Act II alone; Fig. 3A, line 13) fails to activate

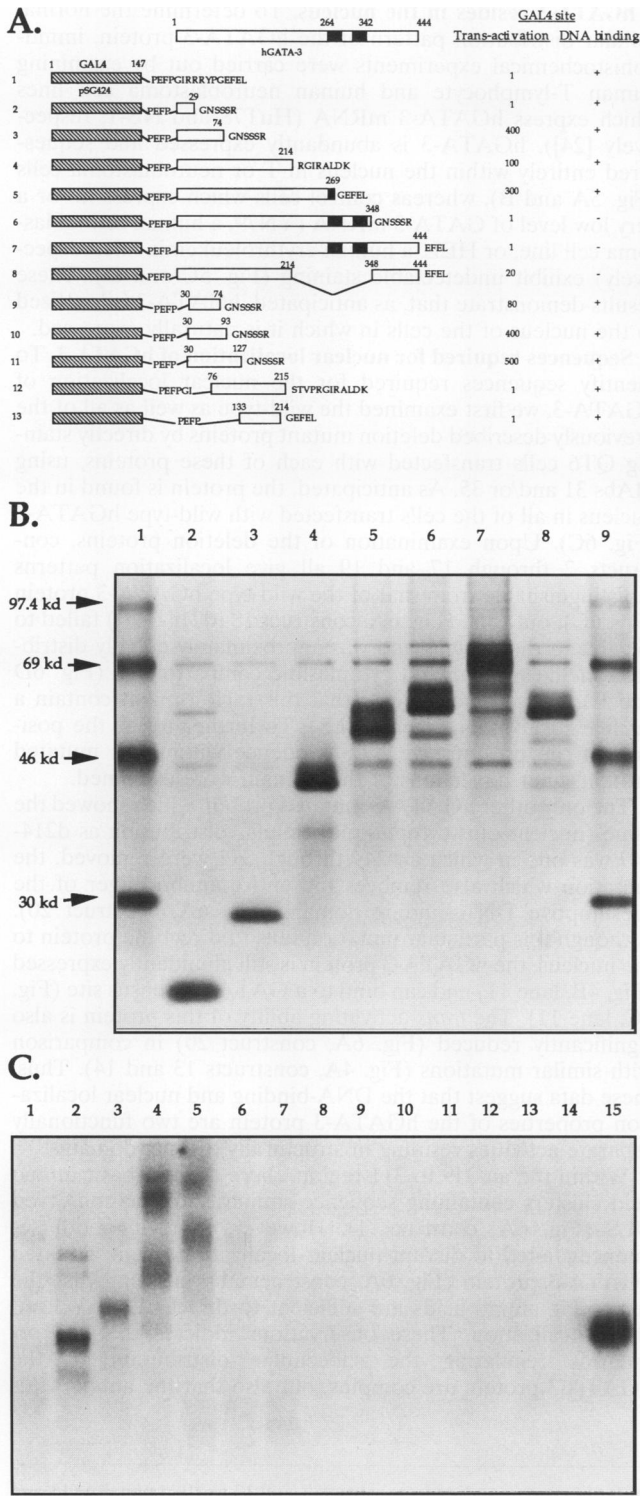

FIG. 3. Fusion chimeras of hGATA-3 and GAL4 reveal the presence of a single independent *trans*-activation domain. (A) GAL4/hGATA-3 fusion constructs and *trans*-activation results. The top line represents the wild-type hGATA-3 protein, followed by individual GAL4/hGATA-3 fusion proteins; the shaded region represents the GAL4 DNA-binding and dimerization domain from aa 1 to 147, joined to different segments of the hGATA-3 protein (as indicated by the amino acid numbers above each construct; also see Table 2). At the right are the summary data for the DNA-binding ability of each of these mutated proteins (C) as well as the reporter gene assay results (relative to *trans* activation by pSG424) as quantified on a PhosphorImager. Each *trans*-activation result is the average of three independent experiments. The calculated standard deviations (percentages) for the transfections are as follows: 1, 0.3; 2, 1.0; 3, 48.2; 4, 15.1; 6, 0.5; 7, 0.8; 9, 8.6; 10, 57.3; 11, 46.2; 12, 1.5; and 13, 1.0. (B) Immunoprecipitation of GAL4/hGATA-3 fusion proteins. QT6 cells transfected with the GAL4/hGATA-3 fusion constructs were labeled with [^{35}S]Met-Cys and immunoprecipitated with MAb 31 (Materials and Methods). Lanes 1 and 9 are protein molecular weight standards, with the sizes indicated on the left. Lanes 2 to 8 reflect the immunoprecipitation results after labeling cells transfected with fusion constructs 2 through 8 (shown in panel A). (C) Fusion protein binding to GAL4 sites. QT6 cells transfected with the GAL4/hGATA-3 chimeras were lysed and assayed by EGMSA as described before. The probe used in lanes 1 to 7 contained five copies of the GAL4-binding sites excised from plasmid pSG5CAT with *Hin*dIII and *Kpn*I, by kinase labeling with [^{32}P-γ]ATP. Because of differences in the sizes of the fusion proteins, the bands migrate at different positions. Since there are five copies of the binding sites in the probe, there are multiple shifted bands in each lane which likely represent different numbers of GATA proteins bound to the probe. Lanes 8 to 15 are the same protein extracts examined using the MαP (GATA)-binding site as probe. Lanes 1 to 7 and 8 to 14 correspond to the protein products of constructs 1 to 7 of panel A, while lane 15 is a positive control (hGATA-3 protein).

GAL4-dependent reporter gene transcription. Similarly, aa 76 to 215 (line 12), which appeared to contain distinct *trans*-activating activities from deletion analysis (Fig. 2A), fail to independently provide a *trans*-activation function when linked to the heterologous DNA-binding domain. Taken together, these data indicate that only aa 30 to 74, or other hGATA-3 coding regions containing Act I, can function to direct transcriptional *trans* activation when fused to a heterologous DNA-binding domain.

Finger II of hGATA-3 is necessary and sufficient for sequence-specific DNA binding. As with all members of this multigene family in vertebrate organisms, the DNA-binding domain of hGATA-3 is composed of two zinc fingers, both of which are highly conserved within the GATA factor family. In mouse and chicken GATA-1, only finger II is required for DNA binding, while finger I appears to be largely dispensable (29, 53). To define the sequences required for DNA binding within hGATA-3, we constructed deletions within the zinc finger domain in which both fingers or each of the two fingers were independently deleted (Fig. 4A). Protein accumulation after transfection was monitored by immunoprecipitation, and all of the deletion mutated proteins were again found to be abundantly expressed (Fig. 4B, lines 3 to 10), even though some appeared to accumulate at a reduced level compared with wild-type hGATA-3 protein.

EGMSA was again performed with extracts recovered from the cells transfected with the finger deletion mutations. When both fingers were deleted, the hGATA-3 protein no longer binds to DNA (Fig. 4A, lines 17 and 18). Proteins in which finger I was deleted can still bind to a consensus GATA site (Fig. 4A, lines 13 and 14). These observations therefore agreed with studies of both the mouse and chicken GATA-1 proteins, in that finger II appears to be the only one required for GATA site-dependent DNA binding. Deletion of finger I reproducibly increased the *trans*-activation ability of the mutated proteins compared with the wild-type hGATA-3 protein (Fig. 4A, lines 13, 14, and 19); it is therefore possible that a repressive domain resides within that region (see Discussion). We conclude that finger II is both sufficient and necessary for DNA binding of the hGATA-3 protein to a GATA consensus binding site.

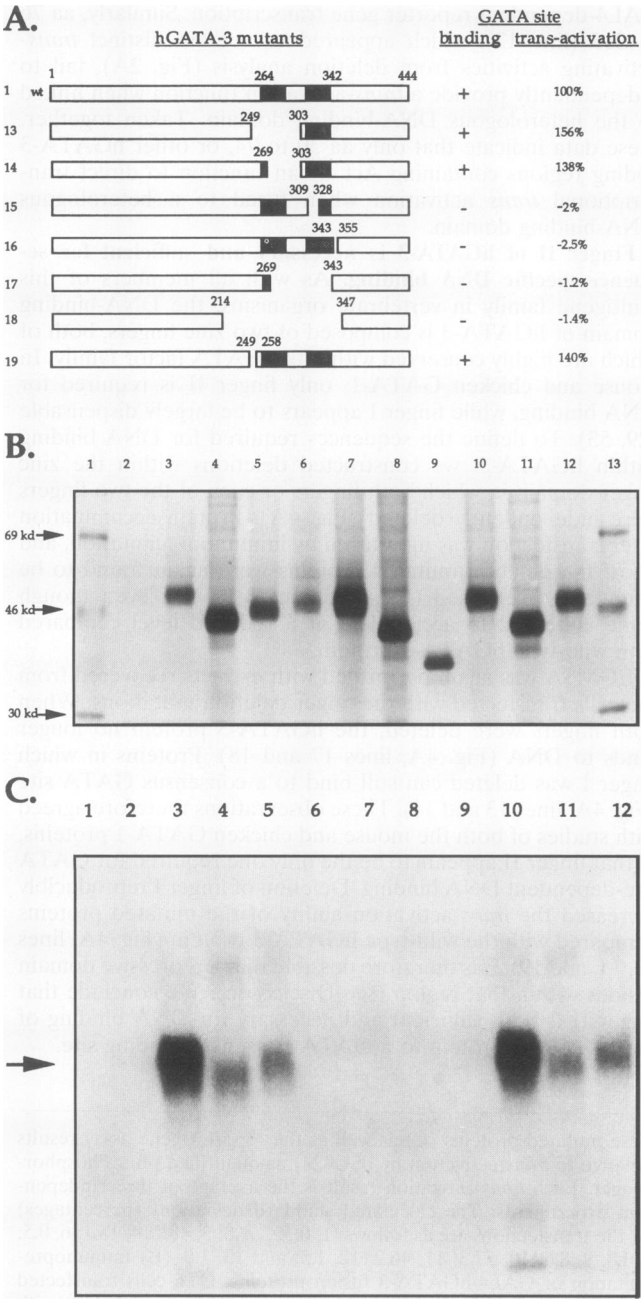

FIG. 4. The DNA-binding domain of hGATA-3 is within finger II. (A) Schematic representation of the deletions within or around the zinc finger region and their DNA-binding and *trans*-activation properties. On the left are the hGATA-3 deletion constructs as described in Fig. 2A. In the middle is the summary of DNA-binding properties; + represents binding and − represents no binding to the TδE4 (GATA-3)-binding site (24) oligonucleotide as shown in panel C. On the right are the *trans*-activation results, representing averages of three independent experiments. The calculated standard deviations of transfections for the above constructs in percentages are as follows: 13, 3.6; 14, 22.8; 15, 1.4; 16, 1.2; 17, 1.5; 18, 1.2; and 19, 7.0. (B) Immunoprecipitation of hGATA-3 deletion mutations. QT6 cells transfected with RSV/hGATA-3 as well as with all of the deletion mutation constructs were labeled with [^{35}S]Met-Cys and immunoprecipitated with MAb 31. Lanes: 1 and 13, molecular weight markers as indicated on the left; 2, labeled extracts from QT6 cells transfected with the hGATA-3 antisense construct (24); 3 to 10, results with use of extracts prepared from QT6 cells transfected with each of the RSV/hGATA-3 deletion mutants, corresponding to constructs 1 and 13 to 19 in panel A; 11 and 12, constructs 20 and 21 described in Fig. 6A. (C) DNA binding. EGMSA experiments were performed as described in the legend to Fig. 2B, and the protein-DNA complex is indicated at the left by the arrow. Lane 1 is radiolabeled oligonucleotide in the absence of added extract; lane 2 includes extracts from QT6 cells transfected with antisense hGATA-3 cDNA. Lanes 3 to 10 represent the gel shift results with extracts prepared from QT6 cells transfected with each of the RSV/hGATA-3 deletion mutations 13 through 19 in panel A. Lanes 11 and 12 correspond to constructs 20 and 21 described in the legend to Fig. 6C.

hGATA-3 resides in the nucleus. To determine the normal cellular distribution pattern of the hGATA-3 protein, immunohistochemical experiments were carried out by examining human T-lymphocyte and human neuroblastoma cell lines which express hGATA-3 mRNA (HuT78 and NB-1, respectively [24]). hGATA-3 is abundantly expressed and sequestered entirely within the nucleus in T or neuroblastoma cells (Fig. 5A and B), whereas control cells which express no or a very low level of GATA-3 mRNA (YN79, a human retinoblastoma cell line, or HEL, a human erythroleukemia line, respectively) exhibit undetectable staining (Fig. 5C and D). These results demonstrate that, as anticipated, hGATA-3 is localized to the nucleus of the cells in which it is naturally expressed.

Sequences required for nuclear localization of hGATA-3. To identify sequences required for the nuclear localization of hGATA-3, we first examined the wild-type as well as all of the previously described deletion mutant proteins by directly staining QT6 cells transfected with each of these proteins, using MAbs 31 and/or 35. As anticipated, the protein is found in the nucleus in all of the cells transfected with wild-type hGATA-3 (Fig. 6C). Upon examination of the deletion proteins, constructs 2 through 17 and 19 all give localization patterns indistinguishable from that of the wild-type hGATA-3 protein (Fig. 6C); only one (Fig. 6A, construct 18 [d214-347]) failed to localize to the nucleus and was approximately equally distributed in the nuclear and cytoplasmic compartments (Fig. 6D and E). These data indicate that this region might contain a nuclear localization signal (NLS). To further refine the position of the presumptive NLS sequence within this mutated protein, finer dissections of this domain were examined.

The only other hGATA-3 mutated protein which showed the same (nuclear plus cytoplasmic) cellular distribution as d214-347 was one in which aa 249 through 311 were removed, the mutation which also removes the entire amino finger of the presumptive DNA-binding domain (Fig. 6A, construct 20). Although this particular mutation fails to direct the protein to the nucleus, the hGATA-3 protein is still abundantly expressed (Fig. 4B, lane 11) and can bind to a GATA consensus site (Fig. 4C, lane 11). The *trans*-activating ability of this protein is also significantly reduced (Fig. 6A, construct 20) in comparison with similar mutations (Fig. 4A, constructs 13 and 14). Thus, these data suggest that the DNA-binding and nuclear localization properties of the hGATA-3 protein are two functionally separate activities residing in structurally distinct domains.

Within the aa 249 to 311 region, there are two basic amino acid clusters containing sequence similarity to the conserved NLS (Fig. 6A, construct 1). However, removing both sequences failed to disrupt nuclear localization of the mutated hGATA-3 protein (Fig. 6A, construct 21), implying that the remaining amino acids are sufficient to direct hGATA-3 nuclear localization. These observations indicate that the sequences regulating the subcellular distribution of the hGATA-3 protein are complex, but also that the amino acids

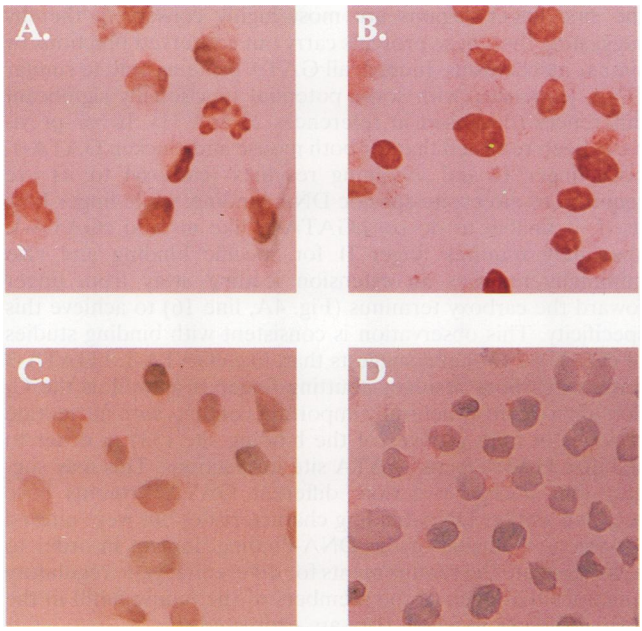

FIG. 5. hGATA-3 is found in the nucleus of both T lymphocytes and neuroblastoma cells. Immunohistochemical analysis of four human cell lines (HuT78, T lymphocyte [A], NB-1, neuroblastoma [B], YN79, retinoblastoma [C], and HEL, erythroleukemia [D]) was carried out as described in Materials and Methods by reaction with MAb 31. The chromogen diaminobenzidine stains antibody-reactive structures brown, while the counterstain methyl green details the positions of the nuclei in all cells. The fixation and photographic procedure used shows only the nuclei in these stained cells.

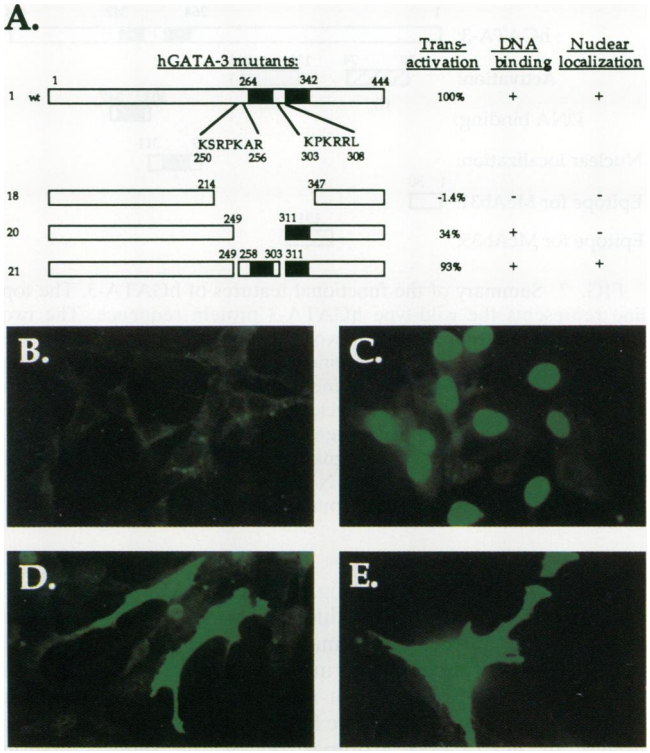

FIG. 6. Immunolocalization of hGATA-3 deletion mutated proteins. (A) Schematic representation of the hGATA-3 deletion mutants. The first line represents wild-type hGATA-3; aa 250 to 256 and 303 to 308 (20, 24) are highlighted; both correspond to the consensus NLS (Discussion). The *trans*-activation results are averages of three independent experiments, the binding results were obtained from Fig. 4B, and the localization results are summarized from Fig. 6B to E and unpublished data. The calculated standard deviations of transfection for constructs 20 and 21 are 1.4 and 11.3%. (B to E) Indirect immunofluorescence staining of hGATA-3 deletion mutated proteins in transfected QT6 cells. QT6 cells transfected with RSV/antisense hGATA-3 (B), with RSV/hGATA-3 (C), or mutated proteins 18 and 20 (D and E, respectively) were fixed in paraformaldehyde (Materials and Methods), treated with MAb 31 or 35, and then treated with fluorescein isothiocyanate-conjugated secondary antibody. When stained with a control primary MAb which recognizes only GST (unpublished observations), the results were identical to those shown in panel B (i.e., only very weak background fluorescence), whereas all other deletion mutation constructs (2 to 17 and 19 shown in Fig. 2 and 4) exhibited the same nuclear staining pattern as shown in panel C.

composing and surrounding finger I comprise much, if not all, of the GATA-3 nuclear localization activity.

DISCUSSION

We report here the characterization of several discrete structural domains which define the properties of a human tissue-restricted transcription factor, hGATA-3. By creating in-frame deletions within the coding region of the hGATA-3 cDNA, we have defined the *trans*-activation domain within the N-terminal region of the protein by transient cotransfection assays, by DNA-binding assays (EGMSA), and by immunoprecipitation with anti-GATA-3 MAbs. These experiments provide substantial evidence that a minimal specific DNA-binding domain resides within finger II (aa 303 through 347) and that a functionally independent *trans*-activation activity is encoded within aa 30 to 74 of this transcription factor. By indirect immunolocalization assays, we show that the hGATA-3 protein is naturally found within the nucleus of cells in which it is expressed (T lymphocytes and neuroblastoma cells) and is also restricted to the nucleus when hGATA-3 cDNA is transfected into cells which do not normally express the protein; aa residues 249 through 311 are (minimally) required for nuclear localization of the hGATA-3 protein. In other experiments (not shown), we have also mapped the epitopes recognized by the two MAbs created here. These data are summarized in Fig. 7.

Sequences required for hGATA-3 *trans* activation are located within the amino-terminal portion of the protein. Using deletion mutations of the hGATA-3 cDNA clone, we initially showed that two regions could be important for the *trans*-activation function of the hGATA-3 protein. Among them, Act I (aa 31 through 59) has a net charge of −4, which is relatively neutral compared with the previously characterized acidic *trans*-activation domains of other proteins such as GCN4, which has 18 acidic and 2 basic amino acids out of 59 in the *trans*-activation domain (17), or VP16, which has 18 acidic out of 78 amino acids (41); both of these are examples of powerful *trans*-activating domains of transcription factors utilizing an "acid blob" motif (43). Act II (aa 132 to 214; Fig. 2A), however, is relatively rich in the amino acids serine (19 of 82) and proline (12 of 82).

When Act I or Act II of hGATA-3 is joined in frame to the GAL4 DNA-binding and dimerization domain, only Act I is capable of directing GAL4-dependent transcription (Fig. 3A), demonstrating that it is a functionally independent *trans*-activation domain of hGATA-3. Act II alone, on the other hand, does not encode an independent *trans*-activation domain, even though its deletion abrogates *trans* activation in the

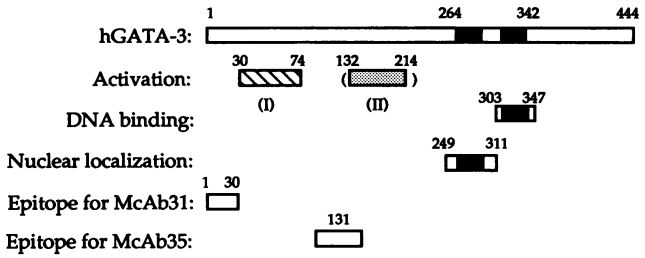

FIG. 7. Summary of the functional features of hGATA-3. The top line represents the wild-type hGATA-3 protein sequence. The two stippled regions represent the two zinc fingers of the protein; the numbers are the first and last amino acids of each functional domain of hGATA-3. This diagram outlines the domains required for *trans* activation, DNA binding, and nuclear localization for the hGATA-3 protein as defined by the experiments described for preceding figures. Also shown are the epitopes recognized by the two antibodies; MAb 31 recognizes an epitope within the N-terminal 30 aa, while the epitope bound by MAb 35 is centered around aa 131.

native hGATA-3 protein (Fig. 2A and 3A). Act II could function in any of several different ways: it may serve to maintain the correct conformation of the folded GATA-3 protein; it may function only in a cooperative manner with a protein containing Act I; or it may exhibit a requirement for acting in concert with a specific DNA-binding domain (the zinc fingers of hGATA-3). Comparison of the *trans*-activation region of hGATA-3 (Act I) with those reported for the mouse and chicken GATA-1 proteins (29, 53) reveals no significant amino acid sequence identity among them, suggesting that these different family members of the GATA factors may interact with an entirely different array of accessory proteins inside the cell and/or may interact with different proteins arranged in a linear array (which in concert form a functional regulatory module like a promoter, enhancer, or silencer [6]) when bound to DNA.

A GAL4 chimera fused to the entire hGATA-3 open reading frame was unable to stimulate GAL4-dependent *trans* activation, whereas a chimera containing the hGATA-3 protein with the zinc finger region deleted was a strong activator (Fig. 3A). Similar observations have previously been reported in examination of fusion proteins consisting of GAL4 linked to AP-2 and to E2-2, CTF-1 joined to Sp1, and GAL4/TEF-1 chimeras; in all of these chimeric transcription factors, the function of a strong activation domain was artifactually masked by the presence of a second functional DNA-binding domain introduced into the fusion protein (15, 18, 32, 51). The molecular basis for this phenomenon is unknown, but in the case of GAL4/TEF-1, the fusion protein can still bind to a GAL4 consensus sequence. In the experiments described here, we cannot totally rule out the possibility that the loss of *trans*-activation ability was due to the inability of the chimeric protein to bind to DNA. We did not detect GAL4 DNA-binding activity (by EGMSA) when a fusion protein contained the whole zinc finger region of hGATA-3 (Fig. 3C), nor did we detect it in the GAL4/hGATA-3 fusion chimera missing the GATA zinc finger domain (data not shown), even though this factor confers low *trans* activation to GAL4-dependent transcription and should therefore bind to DNA in vivo. It is possible the in vitro binding assay used here is not sufficiently sensitive to detect binding when the fusion protein becomes larger than an undefined size.

The minimum DNA-binding domain of hGATA-3 is located in finger II. Among different members of the GATA family, the zinc finger regions are most highly conserved, thereby suggesting that these proteins carry out conserved functions by similar mechanisms. Indeed, all GATA proteins bind to similar DNA sequences, with some potential functionally significant differences (discussed in references 23 and 31). It has previously been reported that for both mouse and chicken GATA-1, both finger II and adjoining residues (carboxyl to it) are required to direct site-specific DNA binding, while finger I, by itself, is unable to do so. hGATA-3, like m- and cGATA-1, absolutely requires finger II for specific binding and also minimally requires an extension leading away from finger toward the carboxy terminus (Fig. 4A, line 16) to achieve this specificity. This observation is consistent with binding studies of cGATA-1 (34) and suggests that, like cGATA-1, hGATA-3 requires carboxy residues abutting finger II as well as the C_4 domain to form structurally important contacts with nucleotide residues in both grooves of the binding site (33) in order to stabilize GATA factor-GATA site interactions. This may suggest that during evolution, different GATA proteins have retained similar DNA-binding characteristics but were altered in structure outside of the DNA-binding domain in order to accommodate the requirements for differential gene regulatory functions that each of the members of the family fulfill in the distinct tissues in which they are expressed.

A defined function for GATA finger I. In this study, we see consistent but relatively modest increases in the activation function of hGATA-3 when the amino finger is removed (Fig. 4A, lines 13 and 14). However, in conjunction with several amino acids upstream of finger I, this structure appears to be required for the protein to localize properly within the nucleus. It has been previously reported that the amino finger is responsible for partial specificity and stability of DNA binding for both the m- and cGATA-1 proteins, and it appears to stabilize this specific DNA/GATA-1 interaction (29, 53).

There are several possibilities to explain the slight increase in the *trans*-activation ability of the hGATA-3 protein containing only the carboxy finger in contrast to the results of m- and cGATA-1 (29, 53). The first and simplest explanation is that finger I encodes a repressive domain, in contrast to the m- and cGATA-1 factors, in which the same domain confers an approximately twofold stimulatory effect (29, 53); however, the precise positional differences of the deletions within these different family members may also account for the contrasting results seen when the GATA-1 and GATA-3 proteins are compared. A second possibility for this disparity is that the reporter gene constructs used were different: here the reporter is transcriptionally directed by three copies of the TCR gene δ enhancer footprint 4, which has a double GATA site (six binding sites in total [24, 39]), whereas one GATA site was used in the construct examining both m- and cGATA-1 domain analysis. A third possibility is that deletion of finger I alters the spacing between different functional domains of the protein through conformational changes which serendipitously decrease the activity (modestly) of GATA-1, while these spacing differences (again, modestly) increase the activity of GATA-3. Since other deletions between Act I and finger II also omit other potentially important functional domains (Fig. 2A, 4A, and 7), we cannot unambiguously conclude that it is the deletion of finger I per se which influences the *trans*-activation activity of any of these proteins.

Detailed dissection of transcription factor proteins has led to the conclusion that functionally important regulatory interactions may lead to cytoplasmic versus nuclear retention of these proteins. For example, there are two signals in the c-Myc protein which are important for localization: one of them acts as an NLS, while the other is also required for transforming

activity (5). Nuclear accumulation of p53 is mediated by several NLSs, each sharing responsibility for nuclear accumulation (42). In another well-characterized example, the NLS is both necessary and sufficient for the nuclear uptake of NF-κB, unless the IκB protein binds to it, masking the NLS, and NF-κB remains in the cytoplasm (3). In other cases, nuclear localization correlates with the cell cycle (50).

We show here that in T lymphocytes and neuroblastoma cells, in which hGATA-3 is normally expressed, or in QT6 cells after transfection of hGATA-3 under the control of a strong constitutive promoter, the hGATA-3 protein is localized entirely within the nucleus. By deletion analysis, we have mapped a region between aa 249 and 311, overlapping finger I, which is required for the nuclear localization of the transfected hGATA-3 protein. Within this aa 249 to 311 region containing and surrounding finger I, there are two short stretches of basic amino acids, KSRPKAR and KPKRRL, both of which are very similar to the consensus NLS (underlined residues) as described previously (4). However, precise excision of either or both segments fails to prevent hGATA-3 transport into the nucleus (Fig. 6A, construct 21, and data not shown). Furthermore, immunofluorescence staining of hGATA-3 harboring smaller deletions (e.g., d269-303 or d249-258; Fig. 4A, constructs 14 and 15, and data not shown) within this 58-aa region showed that each of these also accumulated only in the nucleus, suggesting that multiple signals in the hGATA-3 protein direct nuclear accumulation and that they function independently.

There are other examples of multiple NLSs in a single constitutively expressed nuclear protein which act independently in an additive fashion, where only disruption of all NLS sites abolishes localization (reviewed in reference 44). This has been demonstrated specifically with NS1 of influenza A virus, which has two independent NLSs; only if both are deleted will the protein fail to localize to the nucleus (14). For the tumor suppressor p53, there are three NLSs; deleting any of them results in only partial cytoplasmic localization (42).

It is noteworthy that for both d214-347 and d249-311 hGATA-3 mutated proteins, the immunofluorescence data showed mixed cytoplasmic and nuclear distribution of the protein. This failure in complete compartmentalization could be due to several possibilities, the simplest being that smaller proteins created by deletion can freely diffuse between nucleus and cytoplasm. However, protein size alone cannot be a major determinant for localization of hGATA-3, since proteins smaller than hGATA-3 mutation d249-311 (Fig. 2A, 4A, and 6A) are entirely sequestered within the nucleus (data not shown). A second possibility is that the hGATA-3 protein can enter the nucleus by interacting with other proteins, as is the case for tumor suppressor protein pRB (55). The domains responsible for pRB nuclear localization and protein interaction are both important for successful nuclear translocation, and interestingly, mutations affecting the NLS of pRB retain partial nuclear localization activity due to the ability of mutated pRB to enter the nucleus by associating with other proteins through an intact protein interaction domain. A third possibility is that there are additional NLSs outside aa 249 to 311 of hGATA-3, whose function is at least partially redundant with the activity identified here and which can function in the absence of aa 249 to 311.

It has been reported that the NLSs are incorporated into the DNA-binding domains for a number of nuclear proteins (e.g., Fos [48] and GAL4 [45]). In the case of hGATA-3, the NLS is located within the zinc finger region (finger I) but is separate from the DNA-binding function of the protein (located in finger II). The most conservative interpretation of the data shown here is that the NLS for hGATA-3 is buried around and within the sequence of the amino finger, thereby identifying a new and distinct function for finger I of the GATA proteins.

ACKNOWLEDGMENTS

We appreciate constructive criticisms of the manuscript and provision of the purified GST/cGATA-1, -2, and -3 proteins by our colleague L. J. Ko and provision of the C1300 cells by K. George, and we thank B. Wu for assistance and for providing some of the reagents used in this study.

This work was supported by NIH research grant GM 28896 (J.D.E.).

REFERENCES

1. **Antoniou, M., E. deBoer, G. Habets, and F. Grosveld.** 1988. The human β-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. EMBO J. **7**:377–384.
2. **Arceci, R. J., A. A. J. King, M. C. Simon, S. H. Orkin, and D. B. Wilson.** 1993. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. Mol. Cell. Biol. **13**:2235–2246.
3. **Beg, A. A., S. M. Ruben, R. I. Scheinman, S. Haskill, C. A. Rosen, and A. S. Baldwin, Jr.** 1992. IκB interacts with the nuclear localization sequences of the subunits of NF-κB: a mechanism for cytoplasmic retention. Genes Dev. **6**:1899–1913.
4. **Chelsky, D., R. Ralph, and G. Jonak.** 1989. Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol. Cell. Biol. **9**:2487–2492.
5. **Dang, C. V., and W. M. F. Lee.** 1988. Identification of the human c-Myc protein nuclear translocation signal. Mol. Cell. Biol. **8**:4048–4054.
6. **Engel, J. D., K. M. George, L. J. Ko, J. M. Kornhauser, M. W. Leonard, P. Ting, and M. Yamamoto.** 1991. Transcription factor regulation of hematopoietic lineage cells. Semin. Hematol. **28**: 158–169.
7. **Evans, T., and G. Felsenfeld.** 1989. The erythroid-specific transcription factor Eryf1: a new finger protein. Cell **58**:877–885.
8. **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. USA **85**:5976–5980.
9. **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.
10. **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.
11. **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.

11a. **George, K., and J. D. Engel.** Unpublished data.

11b. **George, K., M. Roth, M. W. Leonard, K. Lieuw, F. Grosveld, and J. D. Engel.** Unpublished data.

12. **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.
13. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
14. **Greenspan, D., P. Palese, and M. Krystal.** 1988. Two nuclear localization signals in the influenza virus NS1 nonstructural protein. J. Virol. **62**:3020–3026.

14a. **Gu, L., B. Andres, L. Ko, Z. Yang, and J. D. Engel.** Unpublished data.

15. **Henthorn, P., M. Kiledjian, and T. Kadesch.** 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κ2 motif. Science **247**:467–470.
16. **Ho, I. C., P. Vorhees, N. Marin, B. K. Oakley, S.-F. Tsai, S. H. Orkin, and J. M. Leiden.** 1991. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor α gene. EMBO J. **10**:1187–1192.

17. **Hope, I. A., and K. Struhl.** 1986. Functional dissection of an eukaryotic transcriptional activator protein, GCN4 of yeast. Cell **46:**885–894.
18. **Hwang, J.-J., P. Chambon, and I. Davidson.** 1993. Characterization of the transcriptional activation function and the DNA binding domain of transcriptional enhancer factor-1. EMBO J. **12:**2337–2348.
19. **Ito, E., T. Toki, H. Ishihara, H. Ohtani, L. Gu, M. Yokoyama, J. D. Engel, and M. Yamamoto.** 1993. Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. Nature (London) **362:**466–469.
20. **Joulin, V., D. Bories, J.-F. Eleouet, M.-C. Labastie, S. Chretien, M.-G. Mattei, and P.-H. Romeo.** 1991. A T-cell specific TCR δ DNA binding protein is a member of the human GATA family. EMBO J. **10:**1809–1816.
21. **Keegan, L., G. Gill, and M. Ptashne.** 1986. Separation of DNA binding from the transcription-activating function of a regulatory protein. Science **231:**699–704.
22. **Kelley, C., H. Blumberg, L. I. Zon, and T. Evans.** 1993. GATA-4: a novel transcription factor expressed in endocardium of the developing heart. Development **118:**817–827.
23. **Ko, L. J., and J. D. Engel.** 1993. DNA-binding specificities of the GATA transcription factor family. Mol. Cell. Biol. **13:**4011–4022.
24. **Ko, L. J., M. Yamamoto, M. W. Leonard, K. M. George, and J. D. Engel.** 1991. Murine and human GATA-3 factors mediate transcription through a *cis*-regulatory element within the human T-cell receptor δ gene enhancer. Mol. Cell. Biol. **11:**2778–2784.
25. **Kornhauser, J. M., M. W. Leonard, M. Yamamoto, J. H. LaVail, K. E. Mayo, and J. D. Engel.** GATA transcription factor expression parallels development of the embryonic chicken optic tectum. Mol. Brain Res., in press.
25a. **Labastie, P., and P.-H. Romeo.** Unpublished data.
26. **Leonard, M. W., K.-C. Lim, and J. D. Engel.** 1993. Expression of the GATA transcription factor family during early erythroid development and differentiation. Development **119:**519–531.
27. **Lindenbaum, M., M. W. Leonard, F. G. Grosveld, and J. D. Engel.** Unpublished data.
27a. **Linzer, D. I. H.** Personal communication.
28. **Lum, S. S. Y., S. Hsu, M. Vaewhongs, and B. Wu.** 1992. The *hsp70* gene CCAAT-binding factor mediates transcriptional activation by the adenovirus E1a protein. Mol. Cell. Biol. **12:**2599–2605.
29. **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:**1886–1898.
30. **Martin, D. I. K., S.-F. Tsai, and S. H. Orkin.** 1989. Increased γ-globin expression in HPFH mediated by an erythroid-specific DNA-binding factor. Nature (London) **338:**435–438.
30a. **Mellon, P.** Personal communication.
31. **Merika, M., and S. H. Orkin.** 1993. DNA-binding specificity of GATA family transcription factors. Mol. Cell. Biol. **13:**3999–4010.
32. **Mermod, N., E. A. O'Neill, T. J. Kelly, and R. Tjian.** 1989. The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell **58:**741–753.
33. **Omichinski, J. G., G. M. Clore, O. Schaad, G. Felsenfeld, C. Trainor, E. Appella, S. J. Stahl, and A. M. Gronenborn.** 1993. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science **261:**438–446.
34. **Omichinski, J. G., C. Trainor, T. Evans, A. M. Gronenborn, G. M. Clore, and G. Felsenfeld.** 1993. A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. Proc. Natl. Acad. Sci. USA **90:**1676–1680.
35. **Orkin, S. H.** 1992. GATA-binding transcription factors in hematopoietic cells. Blood **80:**575–581.
36. **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.
37. **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.
38. **Redondo, J. M., S. Hata, C. Brocklehurst, and M. S. Krangel.** 1990. A T cell-specific enhancer within the human T cell receptor δ locus. Science **247:**1225–1229.
39. **Redondo, J. M., J. L. Pfohl, and M. S. Krangel.** 1991. Identification of an essential site for transcriptional activation within the human T-cell receptor δ enhancer. Mol. Cell. Biol. **11:**5671–5680.
40. **Rosenthal, N.** 1987. Identification of regulatory elements of cloned genes with functional assays. Methods Enzymol. **152:**704–724.
41. **Sadowski, I., J. Ma, S. Triezenberg, and M. Ptashne.** 1988. GAL4-VP16 is an unusually potent transcriptional activator. Nature (London) **335:**563–564.
42. **Shaulsky, G., N. Goldfinger, A. Ben-Se'ev, and V. Rotter.** 1990. Nuclear accumulation of p53 protein is mediated by several nuclear localization signals and plays a role in tumorigenesis. Mol. Cell. Biol. **10:**6565–6577.
43. **Sigler, P. B.** 1988. Acid blobs and negative noodles. Nature (London) **333:**210–212.
44. **Silver, P. A.** 1991. How proteins enter the nucleus. Cell **64:**489–497.
45. **Silver, P. A., L. P. Keegan, and M. Ptashne.** 1984. Amino terminus of the yeast GAL4 gene product is sufficient for nuclear localization. Proc. Natl. Acad. Sci. USA **81:**5951–5955.
46. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides in *Escherichia coli* as fusions with glutathione S-transferase. Gene **67:**31–40.
47. **Trainor, C. D., T. Evans, G. Felsenfeld, and M. S. Boguski.** 1990. Structure and evolution of a human erythroid transcription factor. Nature (London) **343:**92–96.
48. **Tratner, I., and I. M. Verma.** 1991. Identification of a nuclear targeting sequence in the Fos protein. Oncogene **6:**2049–2053.
49. **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 (London) **339:**446–451.
50. **Vris, S., J.-M. Lemaitre, M. Libovici, N. Thierry, and M. Mechali.** 1992. Comparative analysis of the intracellular localization of c-Myc, c-Fos, and replicative proteins during cell cycle progression. Mol. Cell. Biol. **12:**3548–3555.
51. **Williams, T., and R. Tjian.** 1991. Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev. **5:**670–682.
52. **Yamamoto, M., L. J. Ko, M. W. Leonard, H. Beug, S. H. Orkin, and J. D. Engel.** 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. Genes Dev. **4:**1650–1662.
53. **Yang, H.-Y., and T. Evans.** 1992. Distinct roles for the two cGATA-1 finger domains. Mol. Cell. Biol. **12:**4562–4570.
54. **Yang, Z., and J. D. Engel.** 1993. T lymphocyte transcription factor hGATA-3 regulates HIV-1 expression. Nucleic Acids Res. **21:**2831–2836.
55. **Zacksenhaus, E., R. Bremner, R. A. Phillips, and B. A. Gallie.** 1993. A bipartite nuclear localization signal in the retinoblastoma gene product and its importance for biological activity. Mol. Cell. Biol. **13:**4588–4599.
56. **Zon, L. I., C. Mather, S. Burgess, M. E. Bolce, R. M. Harland, and S. H. Orkin.** 1991. Expression of GATA-binding proteins during embryonic development in Xenopus laevis. Proc. Natl. Acad. Sci. USA **88:**10642–10646.
57. **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. USA **87:**668–672.