The Long Terminal Repeat (LTR) of ERV-9 Human Endogenous Retrovirus Binds to NF-Y in the Assembly of an Active LTR Enhancer Complex NF-Y/MZF1/GATA-2*

Xiuping Yu1, Xinguo Zhu1, Wenhui Pi, Jianhua Ling, Lan Ko, Yoshihiko Takeda, and Dorothy Tuan1

From the Department of Biochemistry and Molecular Biology and Institute of Molecular Medicine and Genetics, Medical College of Georgia, Augusta, Georgia 30912

The solitary ERV-9 long terminal repeat (LTR) located upstream of the HS5 site in the human β-globin locus control region exhibits prominent enhancer activity in embryonic and erythroid cells. The LTR enhancer contains 14 tandemly repeated subunits with recurrent CCAAT, GTGGGGA, and GATA motifs. Here we showed that in erythroid K562 cells these DNA motifs bound the following three transcription factors: ubiquitous NF-Y and hematopoietic MZF1 and GATA-2. These factors and their target DNA motifs exhibited a hierarchy of DNA/protein and protein/protein binding affinities: NF-Y/CCAAT > NF-Y/GATA-2 > NF-Y/MZF1 > MZF1/GTGGGGA; GATA-2/GATA. Through protein/protein interactions, NF-Y bound at the CCAAT motif recruited MZF1 and GATA-2, but not Sp1 and GATA-1, and stabilized their binding to the neighboring GTGGGGA and GATA sites to assemble a novel LTR enhancer complex, NF-Y/MZF1/GATA-2. In the LTR-HS5-ep-GFP plasmid integrated into K562 cells, mutation of the CCAAT motif in the LTR enhancer to abolish NF-Y binding inactivated the enhancer, closed down the chromatin structure of the e-globin promoter, and silenced transcription of the green fluorescent protein gene. The results indicated that NF-Y bound at the CCAAT motifs assembled a robust LTR enhancer complex, which could act over the intervening DNA to remodel the chromatin structure and to stimulate the transcription of the downstream gene locus.

The human genome contains a large number of retrotransposons, the L1s, the Alus, and the LTRs of human endogenous retroviruses, which comprise over 40% of the human chromosome DNA (1, 2). The retrotransposons have been suggested to be selfish DNAs that do not serve relevant host functions (3). However, the L1s and Alus contain polymers, the ERV-9 LTRs exhibit an unusual sequence feature, the U3 enhancer and promoter regions (10, 11). The solitary LTRs contain the U3 enhancer and promoter region, the transcribed R region, whose 5’ end marks the initiation site of retroviral RNA synthesis, and the U5 region (12) but no internal gag, pol, and env genes. During primate evolution, the LTRs were self-replicated and stably integrated into various host chromosomal sites (13–15). The ERV-9 LTRs are associated with hematopoietic genes such as the β-like globin genes (15, 16), the selectin genes (GenBankTM accession number AL022146), and the major histocompatibility genes (14). They are also associated with embryonic genes such as the embryonic e-globin gene in the β-like globin gene cluster (15) and the axin gene (16), which is expressed during embryogenesis (17) as well as in adult hematopoietic stem cells (18).

Compared with the LTRs of other families of endogenous retroviruses, the ERV-9 LTRs exhibit an unusual sequence feature, the U3 enhancer regions contain from 5 to 17 tandem repeats (15, 16, 19) comprising four closely related subunits 1, 2, 3, and 4 arranged in the order of (1–2–3–4). The four subunits, each of 40 DNA bases, contain combinations of the following three recurrent DNA sequences (Fig. 1A): (i) ACACAAATCGGCA: the underlined CCAAT motif has been reported to bind the C/EBP family of hematopoietic transcription factors (20) and the ubiquitous transcription factor NF-Y. NF-Y is a trimeric complex composed of NF-YA, -YB, and -YC, in which the YB and YC subunits bind to DNA through their histone fold domains, and the YA subunit, recruited by this dimeric complex, binds through its specific DNA binding domain to the CCAAT motif (21, 22). Expression of NF-YA appears to be developmentally regulated as its level declines from proliferating progenitor cells to terminally differentiated cells during cellular differentiation (23, 24). (ii) TCTGTATCTAGCT: the underlined TATC (GATA) motif has been reported to bind the GATA family of transcription factors, including GATA-1 and -2, which are expressed in hematopoietic progenitor cells and play critical roles in hematopoiesis and globin gene transcription during erythropoiesis (25, 26). (iii) GGTGGGGACGTGGAGA: the underlined G-rich motifs bear high sequence identity to the DNA-binding site of the myeloid transcription factor MZF1, a protein with 13 zinc fingers in which fingers 1–9 and 10–13 bind to the 5’ and 3’ half-sites AGTGGGGA and AGGAGGGGA, respectively, in the target DNA (27–29). MZF1 is expressed in myeloid progenitor cells and has been reported to regulate myelopoiesis (30).

The human β-globin gene locus spans the embryonic e-, the fetal G- and A-y-, and the adult δ- and β-globin genes arranged in the transcriptional order of 5’ e-G-y-A-y-δ-β 3’ in the chromosome. The β-globin locus control region (β-LCR), located 10–60 kb upstream of the globin genes and defined by DNase I–hypersensitive sites HS1–5, plays a pivotal role in regulating transcriptional activation of the far downstream β-like globin genes (31–33). In an effort to define the 5’ border of the β-LCR, we previously cloned and sequenced the DNA further upstream...
of the HS5 site, and we discovered a solitary ERV-9 LTR located 1.5 kb upstream of the HS5 site in the 5′ boundary area of the human β-LCR (15). Subsequent studies showed that this 5′ HS5-ERV-9 LTR possesses prominent enhancer activity in embryonic and erythroid cell lines (16) and in oocytes and progenitor cells, including the erythroid progenitor cells, but not in differentiated somatic tissues in transgenic Zebrafish and humans (34). Unlike the HS2 and HS3 enhancers of the β-LCR, whose enhancer activities are blocked by the HS5 site with reported insulator properties when HS5 is interposed between the enhancer and the cis-linked promoter (35, 36), the ERV-9 LTR enhancer is not blocked by the interposed HS5 site (37).

To characterize further the 5′ HS5-ERV-9 LTR to gain insight into its functional significance, we here identified the transcription factors that bound to and activated the LTR enhancer. Wild type and mutant LTR-GFP plasmids were transfected into leukemic K562 cells that exhibit properties of erythroid progenitor cells and express the embryonic globin gene program (38) and the embryonic teratocarcinoma PA-1 cells derived from primordial oocytes (39). The results showed that base substitutions in the CCAAT, GTGCGGA, and GATA motifs reduced LTR enhancer activity, with the CCAAT mutation generating the largest reduction of up to 98%. Electrophoretic mobility shift assays (EMSA) showed that these core motifs flanked by the respective consensus bases bound three transcription factors, ubiquitous NF-Y and hematopoietic MZF1 and GATA-2. Furthermore, competitive EMSAs and Western blots of proteins that bound to and were eluted from magnetic beads conjugated to the wild type or mutant LTR enhancer probes showed that through protein/protein interactions NF-Y bound at the CCAAT motif recruited MZF1 and GATA-2, but not Sp1 and GATA-1, and stabilized their binding to the neighboring GTGCGGA and GATA sites to assemble a novel LTR enhancer complex, the NF-Y/MZF1/GATA-2. Chromatin immunoprecipitation (ChIP) assays confirmed that these transcription factors and the ERV-9 LTR were associated with one another in vivo. In the LTR-HS5-ep-GFP plasmid integrated into the genome of K562 cells, abolishing NF-Y binding to the LTR enhancer by mutating the CCAAT motif in the LTR enhancer closed down the chromatin structure of the e-globin promoter and silenced the transcription of the GFP gene. The results indicate that the robust LTR enhancer complex assembled by NF-Y could act over the intervening DNA to remodel the chromatin structure and stimulate the transcription of a cis-linked gene locus in embryonic and erythroid progenitor cells.

MATERIALS AND METHODS

Constructs—The wild type (wt) and mutant (4-1)-P-GFP constructs (Fig. 1) were made from pEGFP-C1 vector (Clontech) as follows. The cytomegalovirus enhancer/promoter 5′ of the GFP gene was deleted by Asel and Nhel digestions. The LTR promoter, U3P, generated by PCR with a 5′ cloning site of Asel-BglII and a 3′ cloning site of Nhel was inserted into this cleaved vector to make the U3P-GFP plasmid. To make the (4-1)-P-GFP and the mutant MI-IV-P-GFP plasmids, synthetic oligonucleotides spanning the wt or mutant (4-1) LTR enhancer with the 5′ Asel site and the 3′ Xbal-BglII site were inserted into Asel- and BglIII-cleaved U3P-GFP. The (4-1) LTR enhancer spanned DNA bases 1116–1193, and the LTR promoter P spanned bases 1194–1342 in the ERV-9 LTR (see Fig. 2 in Ref. 15). The mutant enhancer MI-J-MIV contained base substitutions in the CCAAT motif of E1 and GATA and GTGCGGA motif of E4, respectively (see Fig. 1B). To ensure that the GFP gene was enhanced by only the ERV-9 LTR enhancer in these plasmids, the SV40 enhancer/promoter driving the neomycin resistance gene located 3′ of the GFP gene in pEGFP was deleted by Sp1 and SstI digestions. To ensure proper processing of GFP mRNA, the SV40 intron (Promega bulletin 80) was added between EcoRI and SalI sites in the polycloning site region 3′ of the GFP gene. Construction of the E-P-GFP plasmid (Fig. 1B) was described previously (16). The (4-1)-HS5-ep-GFP and MI-HS5-ep-GFP plasmids (Fig. 7) were constructed from the respective (4-1)- and MI-P-GFP plasmids. In these plasmids, the LTR promoter was deleted with Xbal and Agel digestions. Into the cleaved vector, DNA fragments spanning HS5 (base positions 6050–6550, GenBank™ accession number AF064190) and ep amplified by PCR from the LTR-HS5-ep-GFP plasmid (37) by primers tagged with Xbal and Agel sites were inserted. For transfection, all constructs were linearized at an ApaLI site in the vector 300 bases 5′ of the inserts (37).

Cell Culture and Transfection—K562 cells and ovarian teratocarcinoma PA-1 cells (from ATCC) were cultured in RPMI 1640 and Dulbecco’s modified Eagle’s medium, respectively, supplemented with 10% fetal calf serum. For transient transfection, 4 × 10⁶ cells were transfected by electroporation with 20 μg of the linearized plasmid and 3 μg of CMV-β-gal plasmid (Clontech). GFP expression levels of the transfected plasmids were normalized with respect to the level of the β-galactosidase to correct for possible differences in transfection efficiencies as described (16). To obtain K562 cells with integrated plasmids, the linearized plasmids were co-transfected with a pCDNeo plasmid by electroporation (37). The electroporated cells were immediately divided into 3 equal aliquots and cultured separately in medium containing G418 at a final concentration of 400 μg/ml. One month later, the cells were sorted by FACS. The GFP-positive cells were propagated in the same medium, and 10⁶ cells were harvested for FACS analysis and Dnase I hypersensitivity mapping.

Nuclear Extracts and EMSA—Nuclear extracts from 5 × 10⁷ K562 or PA-1 cells were prepared, and EMSA was carried out as described (40). The EMSA reaction (10 μl) contained 2–4 μg of nuclear extract and 40–50 fmol of labeled probe. In competition assays, the unlabeled competitor oligonucleotides were incubated with the nuclear extract for 10 min at 25 °C prior to the addition of the labeled probe. For antibody supershift assays, 2 μl of antibody to NF-YA (Rockland, 200-401-100), GATA-1, and GATA-2 (Santa Cruz Biotechnology, SC-13053x and SC-1235x) or Sp1 (Upstate Technology, 07-124) was added prior to the addition of the probe. The rabbit polyclonal antibody to MZF1 was made by BioSynthesis with a synthetic polypeptide that corresponded to amino acids 444–463 located between the 12th and the 13th zinc fingers in the C terminus of MZF1 (27), a unique region found in the MZF1 family of proteins (GenBank™ accession numbers AAA59898 and AAD55810). The antigenic serum was further purified through protein A-agarose (Amersham Biosciences). The EMSA reactions were resolved in 5% non-denaturing PAGE and analyzed in a PhosphorImager (Storm 840, Amersham Biosciences) as described (40).

Isolation of Proteins Bound to the LTR Enhancer and Western Blots—The biotinylated wild type E2 or mutant E2 (CCAAAT)m (300 pmol, synthesized by Integrated DNA Technologies) was conjugated with 25 mg of Dynabeads M-280 streptavidin (Dynal Biotech) according to vendor protocol. The K562 nuclear extract (12 mg of total protein) was pre-cleared by incubation with poly(dl-dC) as in EMSA for 2 h at 4 °C followed by centrifugation at 12,000 rpm for 10 min at 4 °C, and then incubated with the probe-conjugated magnetic beads (25 mg) on a rotation platform at 4 °C overnight. The protein-bound beads were collected and washed exhaustively with 10 ml each of binding buffer (10 mM Hepes, pH 7.5, 50 mM NaCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 20% glycerol, 2 mM MgCl₂, 0.05% Nonidet P-40) for 7–8 times, followed by the binding buffer containing 0.1 and 0.2 mM NaCl 2 times for each buffer. The bound proteins were eluted from the beads with the binding
buffer containing 0.5 M NaCl. The eluted proteins were detected by Western blots with antibodies of interest using ECL Western blotting detection kit (Amersham Biosciences).

**ChiP and Re-ChiP**—ChiP was performed as described by Johnson and Bresnick (41). The K562 cells were fixed in 1% formaldehyde; nuclei were isolated, sonicated, and pre-cleared with rabbit preimmune serum. Aliquots of the pre-cleared chromatin solution from 5 × 10⁶ cells were either set aside as input or incubated with 5 μg of specific rabbit antibodies to NF-YA, GATA-2, GATA-1 (Santa Cruz Biotechnology, SC-1079x, SC9008x, and SC13053x), MZF1 (BioSynthesis), or pre-immune rabbit IgG on a rotation platform at 4 °C for 3–5 h. After reversal of the cross-links, the DNA was purified from the immuno-complexes and analyzed by isotopic PCR for 29 cycles. The PCR primer pair used to amplify the ERV9-LTR enhancer in quantitative, radioisotopic PCR was forward primer 5′-CTGAGTTTGCTGGGGATGCGAA-3′ and reverse primer 5′-ATACAGACTCCGATTGGTATT-3′. The input DNA template was used as 0.025% of that of the DNA templates in the immunocomplexes. The PCR conditions were 2 pmol of each primer, 0.15 mM dNTPs with 0.2 μCi of [³²P]dCTP (3000 Ci/mmol), and 28 cycles. The PCR products were resolved in a 5% polyacrylamide gel and quantified in a PhosphorImager. In re-ChiP assays, the immuno-complexes with the first antibody were eluted by incubation with 10 mM dithiothreitol at 37 °C for 30 min. The eluates were diluted 50 times with IP dilution buffer and re-immunoprecipitated with a desired second antibody. The DNA was isolated and analyzed as described for ChiP.

**DNase I Hypersensitivity Mapping**—Nuclei were isolated from 1 × 10⁸ K562 cells as described (31). Aliquots (200 μl) of the nuclei were suspended in phosphate-buffered saline, 5 mM MgCl₂, 1 mM phenylmethylsulfonyl fluoride and digested with 0, 2, 4, and 8 units of DNase I (Roche Applied Science). The DNAs purified from the digested nuclei were cleaved with appropriate restriction enzymes, resolved in 1.2% agarose gels, blotted, and hybridized simultaneously with the GFP and HS4 probes as described (31).

**RESULTS**

**Mutations in the CCAAT, GATA, and GTGGGGGA Motifs Reduced LTR Enhancer Activity**—To examine the contribution of the CCAAT, GATA, and GTGGGGGA core motifs to LTR enhancer function, we created wt and mutant (4-1)-P-GFP plasmids that contained base mutations in the respective motifs in the (4-1) enhancer. In these recombinant constructs, the (4-1) subunits were used to represent the LTR promoter. In these recombinant constructs, the (4-1) subunits were used to represent the LTR enhancer, because they span all three motifs to be mutated and are located at the very 3′ end of the LTR enhancer repeats (1–2–3–4)(3–4–1), contiguous with the LTR promoter (Fig. 1A). The wt (4-1) and six mutant enhancers, MI, MII, MIIa, MIIb, MIII, and MIV, were linked to the LTR promoter to create the respective (4-1)-P-GFP and MI-IV-P-GFP plasmids (Fig. 1B). In MI, the CCAAT motif in E1 was mutated to AACCGG (Fig. 1B). In MII, both the GATA (TATC) and the contiguous TAGCTA motifs in E4 were mutated to TAATAGCTCA (Fig. 1B). To determine the relative contribution to LTR enhancer activity of these two sub-motifs, MIla- and MIlb-P-GFP plasmids were created; MIla contained a base substitution only in the GATA moiety, whereas MIlb contained base substitutions only in the TAGCTA moiety (Fig. 1B). In MIII and MIV, the respective 5′-GTGGGGGA and 3′-GTGGGGGA motifs in E4, GTGGGGGACGTGGGAGA (Fig. 1A), were mutated to GAGGAGA (Fig. 1B) to determine their respective contribution to LTR enhancer activity.

The wt (4-1)-P-GFP plasmid and the mutant plasmids together with the E-P-GFP plasmid containing the full-length LTR enhancer spanning the 14 tandem repeats were transiently transfected into erythroid K562 and embryonic teratocarcinoma PA-1 cells. FACS analyses of the GFP fluorescence levels of the transfected cells (Fig. 1C) showed that MI-P-GFP containing the CCAAT mutation exhibited the largest range of reduction in enhancer activity of 50% in K562 and 90% in PA-1 cells. MIII-P-GFP containing base substitutions in the 5′-GTGGGGGA motif exhibited a slightly smaller range of reduction of 50% in K562 cells and 70% in PA-1 cells. In contrast, MIV-P-GFP containing base substitutions in the 3′-GTGGGGGA motif exhibited a much milder reduction of 10–20%, indicating that the 3′-GTGGGGGA motif contributed marginally to LTR enhancer activity. In K562 cells, MII-P-GFP containing mutations in both the TATC (GATA) and the TAGCTA motifs exhibited a reduction in enhancer activity of ~40% in both K562 and PA-1 cells. In K562 cells, the MIla- and MIlb-P-GFP plasmids did not exhibit a significant reduction in activity given the size of the error bars, suggesting that MIla and MIlb enhancers with mutation in either the GATA core or the flanking bases bound the supposed GATA factor almost as well as the wt (4-1) enhancer. However, in PA-1 cells, MIla, MIlb, and MIV exhibited similar reductions in enhancer activity of 40–50%, suggesting that neither MIla and MIlb containing single mutations nor MIV containing double mutations bound the supposed GATA factor as well as the wt (4-1) enhancer. Although the wt (4-1) enhancer showed comparable activity in PA-1 as in K562 cells (Fig. 1C, legend), the full-length LTR enhancer was four times more active in PA-1 cells (Fig. 1C, lane E), apparently because the 14 tandem subunits acted more cooperatively in PA-1 than in K562 cells.

**CCAAT, GTGGGGGA, and GATA Motifs in the LTR Enhancer Bind Transcription Factors NF-Y, MZF1 and GATA-2, but Not Sp1, and GATA-1**—To identify the transcription factors that bind to and activate the wt and the mutant plasmids in K562 and PA-1 cells, we synthesized E1 and E4 probes spanning enhancer subunits 1 and 4, respectively (Fig. 2A), and we used them with K562 and PA-1 nuclear extracts in EMSAs, in competition assays with various competitors, and in antibody supershift assays. Note that in order not to miss potential protein-binding sites in the 3′ junction region, E1 probe extended into the 5′ end of E2 containing the GATA site (Fig. 1A and Fig. 2A).

The E1 probe generated two prominent EMSA bands with the K562 nuclear extract (Fig. 2B, lane 1). The faster migrating band was generated by the GATA (TATC) motif binding to the GATA factor(s), because the (GATA)⁷ competitor of the HSS region (Fig. 2A), previously shown to bind to GATA-1 and weakly to GATA-2 (40), completely obliterated this band (Fig. 2B, lane 4). Furthermore, the antibody to GATA-2 also completely obliterated this band, whereas the antibody to GATA-1 partially obliterated this band (Fig. 2B, lanes 8 and 10). Note that the GATA-2 and GATA-1 antibodies did not supershift but caused the disappearance of the cognate EMSA bands, as observed previously (40). Thus, the results indicate that this fast mobility band was generated by the GATA-2/GATA complex. The partial obliteration of the GATA-2 band by the GATA-1 antibody was not because of cross-interaction of the GATA-1 antibody with GATA-2 (Fig. 4B, lanes 9, 10, and 12). GATA-1 also did not appear to bind directly to the GATA motif in the LTR enhancer. If it did, supershifting GATA-2 with the GATA-2 antibody would have left a residual band of the GATA-1 antibody (Fig. 4B, lane 8). Thus, the partial obliteration of the GATA-2 band by the GATA-1 antibody suggests that GATA-1 might, through protein/protein interaction, directly or indirectly bind to and stabilize the GATA-2/GATA complex.

The slower migrating band of the E1 probe was generated by NF-Y binding to the CCAAT motif. The short E1(GCAAT) competitor spanning only the CCAAT motif in the E1 probe caused the disappearance of
this band (Fig. 2B, lane 5), and the NF-Y antibody supershifted this band (Fig. 2B, lane 9). Interestingly, the CCAAT competitor abolished not only the NF-Y band but also the GATA-2 band (Fig. 2B, lane 5). This indicates that NF-Y bound at the CCAAT competitor could bind to GATA-2 more strongly through the protein/protein interaction than the GATA motif could bind GATA-2 through the DNA/protein interaction. Hence, NF-Y was able to sequester all the GATA-2 from the GATA motif to obliterate the GATA-2/GATA band. The alternative that the E1(CCAAT) competitor containing the sequence CCAATCA (GATT) (see Fig. 2A) could directly bind GATA-2 to sequester all the GATA-2 appeared unlikely, because the CCAATCA competitor when used as a probe in EMSA did not generate the GATA-2 band (Fig. 4C, lane 1). In contrast, the (GATA)7 competitor abolished the GATA-2 band but did not at all affect NF-Y binding to the CCAAT motif (Fig. 2B, lane 4). This indicates that GATA-2 bound at the (GATA)7 competitor could not competitively bind to and sequester NF-Y to abolish the
strong NF-Y/CCAAT complex. The results indicate that the NF-Y/CCAAT complex could competitively bind to and sequester all the GATA-2 in the nuclear extract through protein/protein interactions to prevent GATA-2 binding to the GATA motif. Thus, the order of the binding strengths is as follows: NF-Y/CCAAT > NF-Y/GATA-2 > GATA-2/GATA. This binding hierarchy suggests that in the E1 probe the CCAAT motif first bound NF-Y, which then recruited GATA-2 to the neighboring GATA site to form the E1/NF-Y/GATA-2 complex.

With the PA-1 nuclear extract, the E1 probe generated only a prominent NF-Y band (Fig. 2B, lanes 11 and 15) but no GATA-2 band (Fig. 2B, lanes 11, 16, and 17). This indicates that PA-1 cells did not express GATA-2. The E1 probe thus formed only the E1/NF-Y/GATA-2 complex.

The E4 probe generated three bands with K562 nuclear extract (Fig. 2C, lanes 1 and 8). The strong top band and the weak bottom band contained, respectively, GATA-2 and a GATA x factor of unknown identity, because both bands were obliterated by the (GATA)7 competitor (Fig. 2C, lanes 1 and 10). Only the E4 self-competitor, containing both the E4(MZF) and the GATA motifs, was able to significantly compete out the MZF1 band (Fig. 2C, lane 2). This result indicates that in the E4 probe, the presence of GATA-2 bound at the GATA site greatly strengthened MZF1 binding to the neighboring E4(MZF) site to form a tight E4/MZF1/GATA-2 enhancer complex in K562 cells.

The E1(CCAAT) competitor, in contrast to the E4(MZF) competitor, was able to significantly abolish the E4(MZF) complex and completely abolish the E4/GATA-2 complex (Fig. 2C, lane 11), although the E4 probe did not even contain a CCAAT motif. This indicates that the reduction in MII enhancer activity in K562 (Fig. 1C) was because of the inability of the mutant GATA site to bind GATA-2.

The middle EMSA band was generated by MZF1 binding to the E4(MZF) sequence, GTGGGAGCGTGGAGA that bore high identity to the MZF1 5’ and 3’ half-sites, GTGGGGA and GAGGGGGAA (28), and the E4(MZF) band was supershifted by the MZF1 antibody (Fig. 2D, lane 9). Curiously, this MZF1 band was not at all diminished by the E4(MZF) competitor (Fig. 2C, lane 5), nor was this band significantly competed by the MZF1 5’ and 3’ half-site competitors (Fig. 2C, lanes 9 and 10). Only the E4 self-competitor, containing both the E4(MZF) and the GATA motifs, was able to significantly compete out the MZF1 band (Fig. 2C, lane 2). This result indicates that in the E4 probe, the presence of GATA-2 bound at the GATA site greatly strengthened MZF1 binding to the neighboring E4(MZF) site to form a tight E4/MZF1/GATA-2 enhancer complex in K562 cells.

The E1(CCAAT) competitor, in contrast to the E4(MZF) competitor, was able to significantly abolish the E4(MZF) complex and completely abolish the E4/GATA-2 complex (Fig. 2C, lane 11), although the E4 probe did not even contain a CCAAT motif. This indicates that NF-Y bound at the E1(CCAAT) competitor, through protein/protein interactions, was able to partially sequester MZF1 and completely sequester GATA-2 from their respective DNA-binding sites in E4. Together, the
NF-Y Recruits MZF1 and GATA-2 to the LTR Enhancer

results indicate that the order of binding strengths among the protein/DNA and protein/protein interactions with the E4 probe were NF-Y/CCAAT > NF-Y/GATA-2 > NF-Y/MZF1 > MZF1/GTGGGGA; GATA-2/GATA.

Because PA-1 cells did not express GATA-2 (Fig. 2B, lanes 11 and 17), the E4 probe generated with the PA-1 nuclear extract only a major MZF1 band and three minor bands, including a GATAx band (Fig. 2C, lanes 14, 16, and 17). Again, the MZF1 band was not diminished by the E4(MZF) competitor (Fig. 2C, lane 15) and was abolished only by the E4 probe (Fig. 2C, lane 13). The results indicate that in PA-1 cells MZF1 interacted cooperatively with GATAx to form a tight E4/MZF1/GATAx complex. Hence, the reduction in activity of MII, MIIa, and MIIb mutant enhancers in PA-1 cells (Fig. 1C) could be due to the inability of the mutant GATA site to bind GATAx. Moreover, the absence of GATA-2 in the enhancer complex might enable the 14 sub-units of the full-length LTR enhancer to act more synergistically and thus to exhibit four times higher activity in PA-1 than in K562 cells (Fig. 1C).

To confirm that in the E4 probe the E4(MZF) sequence, GTGGGG-GAGGTGGGGA, indeed bound MZF1, E4(MZF) was used as a probe with K562 nuclear extract and various competitors (Fig. 2D). E4(MZF) probe generated three very faint bands even after prolonged exposure of the autoradiogram (Fig. 2D, lane 1), in contrast to the strong EMSA bands generated by the E4 probe spanning both the MZF and the GATA motifs (Fig. 2C). This indicates that the E4(MZF) motif bound weakly to MZF1 and required GATA-2 bound at the neighboring GATA site in E4 to form the strong E4/MZF1/GATA-2 band.

The top two E4(MZF) bands were generated mainly by Sp1, because the Sp1 antibody decreased the intensity of this band (Fig. 2D, lane 8). The bottom band was generated by MZF1 binding to the E4(MZF) 5′ motif, GTGGGGGA, because the E4(MZF) 5′ competitor abolished this band, and the MZF1 antibody supershifted this band (Fig. 2D, lanes 5 and 9), whereas the E4(MZF) 3′ competitor, GTGGGGA with a G → A substitution, did not abolish the MZF1 band (Fig. 2D, lane 6) and therefore did not efficiently bind MZF1. It is thus possible that the E4 (MZF) 5′ site, GTGGGGGA, bound MZF1, whereas the 3′ site, GTGGGGA, contributed to the strength of this binding. Furthermore, the E4(MZF) 5′ site competitor, GAGGGGA with two G → A mutations in the 5′ site, did not decrease the intensity of the MZF1 band (Fig. 2D, lane 7) and was therefore unable to bind MZF1. Because this mutant 5′ site was identical to that in the MIII enhancer (Fig. 1B), the 50% reduction in enhancer activity of the MIII-P-GFP plasmid (Fig. 1C) was because of the inability of the mutant E4(MZF) 5′ site to bind MZF1. In contrast, when the much weaker E4(MZF) 3′ site was mutated to the same GAGGGGA sequence in the MIV-P-GFP plasmid (Fig. 1B), enhancer activity was reduced by only ~15% (Fig. 1C).

It is of interest to note that although the E4(MZF) probe did not contain a CCAAT or a GATA motif, the E1(CCAAT) and E4(GATA) competitors were able to diminish the MZF1 band but not the Sp1 bands (Fig. 2D, lanes 3 and 4). This result again indicates that NF-Y and GATA-2 bound at the respective CCAAT and GATA motifs could competitively bind MZF1 through protein/protein interactions to sequester MZF1 from the E4(MZF) probe.

Together, the results indicate that in K562 cells MZF1 and GATA-2 bound weakly to the respective GTGGGGGA and GATA motifs, but they bound cooperatively to the E4 probe spanning both motifs to form a strong E4/MZF1/GATA-2 complex. However, NF-Y bound strongly at the CCAAT motif in E1. Through protein/protein interactions, it could recruit and stabilize binding of GATA-2 and MZF1 to the neighboring GATA and GTGGGGGA motifs in E4 to assemble the enhancer complex (E1/NF-Y/GATA-2)(E4/MZF1/GATA-2). In PA-1 cells, the enhancer complex appeared to be (E1/NF-Y)(E4/MZF1/GATAx).

NF-Y Bound at the CCAAT Motif Recruits and Stabilizes GATA-2 Binding to the Neighboring GATA-2 Site—The EMSA results indicate that NF-Y bound at the CCAAT motifs interacted with GATA-2 and MZF1 and stabilized their binding to the neighboring GATA and GTGGGGGA sites. To confirm this property of NF-Y, we synthesized the E2 probe containing both the GATA and the CCAAT motifs and a mutant E2 probe, E2(CCAAT)m, containing a mutant CCAAT motif, AACCG, but a wild type GATA-2 motif (see Fig. 3A). As expected, the E2 probe generated with the K562 nuclear extract the NF-Y and GATA-2 bands (Fig. 3B, lanes 1–7). The E2(CCAAT)m probe at the 3′ end of E2 (see Fig. 1A) could be a potential binding site for the EKLF family of transcription factors important in the transcriptional regulation of the globin genes (42). However, neither the E2(CCACT) nor the LTR promoter P(CACCC) competitor changed the intensity or the pattern of the EMSA bands (Fig. 3B, lanes 8 and 9), therefore, the E2(CCACT) motif did not bind any nuclear proteins in K562 cells.

Interestingly, the E2(CCAAT)m probe, containing the mutated CCAAT motif but a normal GATA motif, generated no NF-Y band and
NF-Y Recruits MZF1 and GATA-2 to the LTR Enhancer

also no clearly discernible GATA-2 band (Fig. 3B, lane 11). This result indicates that the GATA site in E2 was a weak site. By itself, it was unable to bind significant amounts of GATA-2 and required NF-Y bound at the neighboring CCAAT motif to recruit and stabilize GATA-2 binding to generate the strong GATA-2/E2 band (Fig. 3B, lane 4). This interpretation suggested that the GATA-2/E2 band should contain both the NF-Y/CCAAT and GATA-2/GATA complexes, which, however, raised an apparent paradox. The GATA-2/E2 band in the EMSA gel clearly contained only the GATA-2/GATA complex without the detectable presence of the NF-Y/CCAAT complex, because the NF-Y antibody did not abolish or supershift the GATA-2/E1 or the GATA-2/E2 band (Fig. 2B, lane 9; Fig. 3B, lane 1).

One possible explanation for this apparent paradox is that the steady state of the DNA/protein interactions detected in EMSA was not a static state but represented a dynamic equilibrium of protein/DNA and protein/protein interactions (43, 44). Thus, in this network of dynamic association and dissociation reactions, NF-Y was first recruited to the CCAAT motif in the E2 probe; the NF-Y/CCAAT complex then served as a scaffold to recruit GATA-2 to the adjacent weak GATA-2 site to form a stable GATA-2/GATA complex, even after NF-Y was subsequently dissociated from the neighboring CCAAT site to bind to other unoccupied CCAAT sites in the free E2 probes. Because the free probe was present in a large molecular excess over the transcription factors in the EMSA reaction, the free E2 probe was able to drive the dynamic association and dissociation reactions to an equilibrium, where the occupied probes contained predominantly only one bound protein per probe to generate either the GATA-2/E2 or the NF-Y/E2 band observed in the EMSA gels.

The E2(CCAAT)m probe unable to bind NF-Y was identical in sequence to the mutant MI enhancer (Fig. 1B). Hence, the reduced enhancer activity of the MI-P-GFP plasmid (Fig. 1C) was due primarily to the inability of the MI enhancer to bind NF-Y in K562 and PA-1 cells. Because NF-Y bound at the CCAAT motif in the LTR enhancer could recruit and stabilize GATA-2 binding to the neighboring GATA site, the MIIa and MIIb enhancers with mutated GATA core or flanking bases (Fig. 1B) could apparently still bind GATA-2 and therefore showed minimal reduction in enhancer activity in K562 cells (Fig. 1C, lanes MIIa and MIIb).

Consensus 5'- and 3'-Flanking Bases Determine Preferential GATA-2 Binding to the AGATA Core—Analysis of the GATA-2-binding sites in E1, E2, and E4 probes (supplemental Table S1) showed that these sites all contained the AGATA core, which is present also in the GATA-1-binding sites (45, 46). The existence of a common core motif for different GATA factors suggests that bases flanking the core determined preferential binding to GATA-1 or GATA-2. To identify the consensus flanking bases, we tabulated the GATA probes and competitors used in EMSAs and deduced a consensus GATA-2-binding site, GGAGCTAGATA (supplemental Table S1, line 9 GATA-2 site). To test whether the consensus flanking sequences conferred preferential GATA-2 binding, we synthesized a mutant GATA probe, syn. GATA-2m, that contained the nonconsensus 5’- and 3’-bases flanking the AGATA core (Fig. 3A; supplemental Table S1, line 7 GATA-2 site) and used it as a competitor in EMSA with the E2(GATA) probe containing the consensus 5’- and 3’-flanking bases (Fig. 3A; supplemental Table S1, line 1 GATA-2 site). With K562 nuclear extract, the E2(GATA) probe generated two faint bands as observed earlier (Fig. 3B, lane 11). These two bands were abolished by the unlabeled self-competitive at concentrations 100–200 times that of the probe (Fig. 3C, left panel). In contrast, the syn GATA-2m competitor did not diminish the EMSA bands at concentrations 100–200 times that of the probe (Fig. 3C, right panel). The results indicate that the syn GATA-2m sequence with nonconsensus 5’- and 3’-flanking bases did not bind GATA-2 and therefore did not act as an efficient competitor. The results indicate that the consensus 5’- and 3’-flanking bases conferred on the AGATA core selective binding to GATA-2 over GATA-1.

Hierarchy of Binding Strengths in the LTR Enhancer Complex—The competitive EMSAs showed that NF-Y bound at the CCAAT motif was highly competitive; through protein/protein interactions, NF-Y could competitively bind to and sequester GATA-2 and MZF1 to abolish the GATA-2/GATA and MZF1/GTGGGGA complexes (Fig. 2, B, lane 5, C, lane 11, and D, lane 3). NF-Y was also a selective competitor, as it did not bind competitively to Sp1 to abolish the Sp1/E4(MZF) complex (Fig. 2D, lane 3). In comparison, MZF1 bound at the GTGGGGA motif was less competitive, as it was unable to competitively bind NF-Y and GATA-2 to dissociate the NF-Y/CCAAT and the GATA-2/GATA complexes in the E1 probe (Fig. 2B, lane 6). However, it was able to partially sequester GATA-2 to partially abolish the GATA-2/GATA complex in the E4 probe (Fig. 2C, lane 10). Similarly, GATA-2 bound at the GATA motif was not competitive; it was unable to bind to NF-Y to abolish the NF-Y/CCAAT complex (Fig. 2B, lane 4) or to MZF1 to completely abolish the MZF1/GTGGGGA complex (Fig. 2, C, lane 3, and D, lane 4). Furthermore, MZF1 and GATA-2 separately bound to their individual cognate motifs only very weakly (Fig. 2D, lane 1, and Fig. 3, B, lane 11, and C); however, they interacted cooperatively to bind strongly to the E4 probe spanning both the MZF and the GATA motifs (Fig. 2C, lanes 1–5). Thus, the binding strengths among the proteins and their cognate DNA motifs were NF-Y/CCAAT > NF-Y/GATA-2 > NF-Y/MZF1 > GATA-2/MZF1 > MZF1/GTGGGGA; GATA-2/GATA.

The LTR Promoter Binds NF-Y, Sp1, and GATA-2—The results presented above indicate that NF-Y bound at the CCAAT motif served as a central scaffold to recruit GATA-2 and MZF1 to assemble the LTR enhancer complex. However, transfection experiments showed that the MI-P-GFP plasmid containing the mutated CCAAT motif exhibited only a 50% reduction in enhancer activity (Fig. 1C). This could be due to the presence of an additional CCAAT motif in the LTR promoter (P) in MI-P-GFP plasmid (Fig. 1A), which could bind NF-Y to mitigate the otherwise more deleterious effect of the CCAAT mutation in the MI enhancer. The LTR promoter also contained the GATA and the GTGGGGA motifs (Fig. 1A), which could bind GATA-2 and MZF1, to mitigate the effects of the mutations in the MII and MIII enhancers (Fig. 1C).

To investigate the proteins that bound to the CCAAT, GTGGGGA, and GATA motifs in the LTR promoter, we carried out EMSA using the LTR promoter probe (Fig. 4A, P probe). With the K562 nuclear extract, the P probe generated three major bands (Fig. 4B, lane 1). The fastest migrating band was produced by GATA-2 binding to the GATA motif (Fig. 4B, lanes 5, 9, and 10). The middle band was produced by NF-Y binding to the CCAAT motif (Fig. 4B, lanes 3 and 8). The slowest migrating band was produced by binding of Sp1, but not MZF1, to the GTGGGGA sequence containing the core motif GGGTG (CACCC), because the band was supershifted by Sp1 antibody but not by MZF1 antibody (Fig. 4B, lanes 7 and 11). Note that in the competitive EMSA, the NF-Y/CCAAT competitor almost completely abolished the GATA-2 band but did not appreciably affect the Sp1 band (Fig. 4B, lane 3). Conversely, the P(CACCC+GATA) competitor spanning the Sp1 and GATA-2 sites in the LTR promoter abolished the Sp1 and GATA-2 bands but not at all the NF-Y band (Fig. 4B, lane 4). These results confirmed earlier EMSAs with the LTR enhancer probes (Fig. 2) that NF-Y interacted with GATA-2 but not with Sp1. The ACCAC(GTGGT) motif in the LTR promoter was a
whether the flanking bases were important in preferential NF-Y binding to CAGCANNCTG (supplemental Table S1, line 7 NF-Y site). To test - and 3'-flanking bases, GTNAATNCA and GTGGGGA motifs in the LTR promoter, which contained six base mutations from the consensus 5'-flanking bases (Fig. 4A). However, the E2(CCAAT)m probe did not efficiently bind GATA-2 (Fig. 5B). We then used the probe-conjugated beads to isolate proteins from the K562 nuclear extract that bound to the probes. Resolution of the isolated proteins in PAGE showed that the E2(CCAAT)m probe, compared with the wt E2 probe, bound poorly to 4–5 proteins in the molecular mass range of 40–60 kDa (Fig. 5A, bands marked with dots). Western blot of the gel with the NF-YA antibody showed that the mutant E2(CCAAT)m probe did not bind NF-Y; consequently, the normal GATA motif in the E2(CCAAT)m probe did not efficiently bind GATA-2 (Fig. 5B, NF-YA and GATA-2 panels). Although the E2 probe did not contain the MZF1 motif, NF-Y and GATA-2 bound at the E2 probe, through protein/protein interactions, were able to bind MZF1, whereas the E2(CCAAT)m probe containing the mutated CCAAT motif showed greatly diminished but a residual ability to bind MZF1 (Fig. 5B, MZF1 panel). In contrast, the control blot with the GATA-1 antibody showed that neither the E2 nor the E2(CCAAT)m probe bound GATA-1 (Fig. 5, GATA-1 panel). These results provided further evidence for the hierarchy of DNA/protein and protein/protein binding affinities that NF-Y bound at the CCAAT motif could recruit and stabilize binding of GATA-2 and MZF1 to their cognate DNA motifs in the LTR enhancer.

NF-Y Recruits MZF1 and GATA-2 to the LTR Enhancer

NF-Y Assembles the LTR Enhancer Complex, NF-Y/GATA-2/MZF1, through Protein/Protein Interactions—The EMSAs (Figs. 2–4) indicate that NF-Y bound at the CCAAT motif was able to recruit GATA-2 and MZF1 and to stabilize their binding to the cognate GATA and GTGGGGA motifs in the LTR enhancer through protein/protein interactions. To confirm this finding, we conjugated to magnetic beads the E2 enhancer subunit containing both the CCAAT and the GATA motifs or the mutant E2(CCAAT)m containing the GATA motif but a mutated CCAAT motif unable to bind NF-Y (see Fig. 3A). We then used the probe-conjugated beads to isolate proteins from the K562 nuclear extract that bound to the probes. Resolution of the isolated proteins in PAGE showed that the E2(CCAAT)m probe, compared with the wt E2 probe, bound poorly to 4–5 proteins in the molecular mass range of 40–60 kDa (Fig. 5A, bands marked with dots). Western blot of the gel with the NF-YA antibody showed that the mutant E2(CCAAT)m probe did not bind NF-Y; consequently, the normal GATA motif in the E2(CCAAT)m probe did not efficiently bind GATA-2 (Fig. 5B, NF-YA and GATA-2 panels). Although the E2 probe did not contain the MZF1 motif, NF-Y and GATA-2 bound at the E2 probe, through protein/protein interactions, were able to bind MZF1, whereas the E2(CCAAT)m probe containing the mutated CCAAT motif showed greatly diminished but a residual ability to bind MZF1 (Fig. 5B, MZF1 panel). In contrast, the control blot with the GATA-1 antibody showed that neither the E2 nor the E2(CCAAT)m probe bound GATA-1 (Fig. 5, GATA-1 panel). These results provided further evidence for the hierarchy of DNA/protein and protein/protein binding affinities that NF-Y bound at the CCAAT motif could recruit and stabilize binding of GATA-2 and MZF1 to their cognate DNA motifs in the LTR enhancer.

In Vivo Interactions of NF-Y, GATA-2, and MZF1 with the ERV-9 LTR Enhancer—We next used ChIP to determine whether these protein factors also bind in vivo to the ERV-9 LTR enhancer in K562 cells. In the PCR step of the ChIP assay, the forward primer was located in the U3 region immediately upstream of the LTR enhancer, and the reverse primer was located in subunit 1 of the second (1–2–3–4) enhancer repeat (see Fig. 6A). However, the reverse primer annealed also to two other subunit 1s in the LTR enhancer but not to the last subunit 1 next to the LTR promoter, which contained six base mutations from the sequence of the reverse primer (Fig. 6A). Hence, this primer pair amplified three PCR bands, the anticipated band of 232 bp and two additional bands of 72 and 392 bp (Fig. 6A, lanes 1–7). Comparison of the intensities of the PCR bands showed that the LTR enhancer was associated in vivo with NF-Y, GATA-2, and MZF1 but not with NF-E2 and GATA-1 (Fig. 6A, right panel). Because the human genome contains 3000–4000 copies of the ERV-9 LTRs, the PCR bands were amplified not only from the β-globin LTR but also from other ERV-9 LTRs, which contained identical U3 enhancer sequences (16, 19), were associated with actively

FIGURE 4. EMSA of the LTR promoter and the e-globin promoter probes with K562 nuclear extract. A, DNA sequences of the probes and the competitors. P, the LTR promoter; P(CCAAT), the CCAAT site in the LTR promoter; eP(CCAAT), the CCAAT site in the e-globin promoter. B and C, EMSA gels of the LTR promoter probe and the P(CCAAT) and the eP(CCAAT) probes.

FIGURE 5. Western blots of K562 nuclear proteins bound to the E2- and E2(CCAAT)m. A, PAGE of equal amounts of the K562 nuclear proteins eluted from magnetic beads conjugated to either the E2- or the E2(CCAAT)m probes. M, size markers in kilodaltons as marked on the left margin. Bands were visualized with silver staining. Dots mark proteins detectable in the E2 lane but not in the E2(CCAAT)m lane. B, Western blots of similarly run PAGE with antibodies to NF-YA, GATA-2, MZF1, and GATA-1 respectively. Numbers in the right margins: sizes (kDa) of the detected protein bands.
NF-Y Recruits MZF1 and GATA-2 to the LTR Enhancer

To investigate whether NF-Y bound to the ERV-9 LTR enhancer was associated in vivo with GATA-2 and MZF1, we carried out the re-ChIP assays. The results showed that the LTR enhancer pulled down by the NF-YA antibody in the first round of ChIP (1st ChIP) could be pulled down again by the antibody of either NF-Y or GATA-2 in a subsequent ChIP (2nd ChIP) (Fig. 6B, lane 4). Conversely, the LTR enhancer pulled down by the GATA-2 or the MZF1 antibody in the 1st ChIP could again be pulled down by the antibody of NF-Y or GATA-2 (Fig. 6B, lanes 3 and 5). The re-ChIP results thus indicate that transcription factors NF-Y, GATA-2, and MZF1 were associated with the ERV-9 LTR enhancer in vivo in K562 cells.

NF-Y Bound at the CCAAT Motif in the LTR Enhancer Remodels Chromatin Structure of the Downstream e-Globin Promoter—NF-Y bound at the CCAAT motif in the promoter has been reported to prevent the formation of nucleosomes around the CCAAT site (49), and can thus remodel the chromatin structure of the resident promoter (50, 51). These findings raised the question of whether NF-Y bound at the CCAAT site in the LTR enhancer could remodel the chromatin structure of a cis-linked promoter over a distance. To address this question, we constructed the wild type (4-1)-HS5-globin promoter, which, unlike the LTR promoter, contained the CCAAT motif unable to bind NF-Y (Fig. 4C). The HS5 site of 500 DNA bases (37) is located naturally downstream of the ERV-9 LTR in the genome and contains no CCAAT motifs; it was inserted between the LTR enhancer and the e-promoter to provide the intervening distance to resolve the promoter site from the enhancer site for DNase I hypersensitivity mapping by Southern blots. The control and the test plasmids were separately integrated into K562 cells.

Southern blots of the integrated plasmids showed that in the (4-1)-HS5-e-GFP plasmid containing the wt LTR enhancer, the e-globin promoter exhibited a DNase I-hypersensitive site (Fig. 7B, left panel). The result indicates that the e-globin promoter in this plasmid existed in an open and accessible chromatin structure. Accordingly, FACS analysis showed that the GFP gene was active and produced a high GFP fluorescence level of over 5000 (Fig. 7B, right panel). On the other hand, in MI-HS5-e-GFP plasmid, which contained the mutant CCAAT motif unable to bind NF-Y, the DNase I-hypersensitive site in the e-globin promoter was not detected (Fig. 7B, right panel). The inability to detect the promoter hypersensitive site was not because of under-digestion of.

**FIGURE 6.** ChIP and re-ChIP assays. A, top panel, map of the ERV-9 LTR. E, the LTR enhancer region. Horizontal line with triangular ends, anticipated PCR product amplified by the PCR primers; horizontal lines above and below, additional PCR products amplified by the PCR primer pair; numbers, sizes in bp of the respective PCR products. Bottom left panel, a representative autoradiogram of the PCR products. Lanes 1–7, PCR bands generated from the chromatin in the input and in test samples pulled down by the antibodies in nonimmunized rabbit serum (IgG) and the antibodies to the respective transcription factors as marked. Numbers in the left margin, sizes in bases of the PCR products. Bottom right panel, relative intensities of the 72-bp PCR bands; the intensity of the input band was set at 1 and served as the reference for comparison. Values are averages of two independent experiments. B, re-ChIP assays. In the 1st ChIP, the chromatin was pulled down with IgG, GATA-2, NF-YA, and MZF1 antibodies as marked horizontally at top of the gels. In the 2nd ChIP, the pulled down chromatin was subsequently interacted with NF-YA, GATA-2, or IgG antibodies as marked vertically in the left margin. PCR products shown are the 72-bp bands in the PAGE.
NF-Y Recruits MZF1 and GATA-2 to the LTR Enhancer

DISCUSSION

The solitary ERV-9 LTR located upstream of the β-globin LCR contains in the enhancer region 14 tandem repeats of four closely related subunits 1, 2, 3, and 4 arranged in the order (1–2–3–4)–(4–1). In this study, we showed that these enhancer subunits contain three recurrent sequences, ACACCAATCGACA, GGTGGGGACGTGGAGA, and AGCTAGATACAGAGT, with the underlined core motifs and conserved flanking bases that bound both in vitro and in vivo the ubiquitous transcription factor NF-Y and the hematopoietic transcription factors MZF1 and GATA-2. As all three protein factors are expressed preferentially in progenitor cells (23, 25, 30), their binding to the LTR enhancer may provide a molecular basis for the earlier observation that the ERV-9 LTR enhancer is active in progenitor cells but not in differentiated somatic cells in transgenic Zebrafish and humans (34). Consistent with the finding that the ERV-9 LTR enhancer is active also in the primordial oocytes of transgenic Zebrafish and humans (34), the ERV-9 LTR enhancer bound to and was activated by NF-Y, MZF1, and a GATAx factor in ovarian PA-1 cells derived from primordial oocytes (39).

In assembling the LTR enhancer complex, the protein/DNA and protein/protein interactions exhibited the following hierarchy of binding affinities: NF-Y/CCAAT > NF-Y/GATA-2 > NF-Y/MZF1 > MZF1/GATA-2 > MZF1/GTGGGGA; GATA-2/GATA. The strong affinity between NF-Y and the CCAAT motif, at the top of the binding hierarchy, is consistent with earlier reports that NF-Y-A in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif, is consistent with earlier reports that NF-YA in the trimeric complex through a specific DNA binding domain binds to the CCAAT motif. The nearly complete loss in enhancer activity of the mutant MI plasmid expressed GFP at a level only 2% that of the wt (4–1) plasmid (Fig. 7C, right bar).

The nearly complete loss in enhancer activity of the mutant MI enhancer due to the CCAAT mutation indicates that NF-Y bound at the CCAAT motif interacted in vivo with MZF1 and GATA-2 to assemble an active LTR enhancer complex, NF-Y/MZF1/GATA-2, which could act over the distance of the intervening HS5 site to remodel the chromatin structure of the β-globin promoter and activate transcription of the GFP gene.

In erythroid K562 cells, MZF1 and GATA-2 appeared to be the preferred protein partners for NF-Y. Our results showed that NF-Y preferentially interacted with GATA-2 but not with GATA-1, even though GATA-2 is less abundant than GATA-1 in K562 cells (53). On the other hand, NF-Y and GATA-1 bound at the respective GATA and CCAAT sites in the promoter of the erythroid Gfi-1B gene have been reported to cooperatively activate the Gfi-1B promoter in K562 cells (54). However, this GATA site bound very weakly to GATA-1 but strongly to an unidentified protein (54). Because direct protein/protein interaction between NF-Y and GATA-1 was not shown in this study, the possibility cannot be excluded that the observed cooperativity between NF-Y and GATA-1 in K562 cells was mediated indirectly through their interactions with an unidentified third protein and not by direct NF-Y/GATA-1 interaction.

In K562 cells, NF-Y did not interact with Sp1. This finding appears to contradict earlier reports that NF-Y bound at the CCAAT motif cooperated with Sp1 bound at the neighboring Sp1 site to activate the resident promoters (22, 55–57). Those studies were, however, carried out with nuclear extracts of nonhematopoietic cells, which apparently did not express MZF1 and GATA-2. Indeed, in the EMSA carried out with the nuclear extract of Friend erythroid leukemia cells, NF-Y bound at the CCAAT motif in the ferritin heavy chain promoter did not recruit Sp1 to the consensus Sp1 sites flanking the CCAAT motif (58). These findings indicate that in hematopoietic cells, NF-Y interacted preferentially with MZF1 and GATA-2 but not with Sp1. However, in nonhematopoietic cells that did not express MZF1 and GATA-2, NF-Y then interacted with the ubiquitously expressed Sp1.

The strong affinity between NF-Y and the CCAAT motif, at the top of the binding hierarchy, indicates that in the enhancer repeats of ERV-9 LTR, (1–2–3–4), the CCAAT motifs in subunits 1, 2, and 3 initially bound NF-Y, which then recruited MZF1 and GATA-2 to the neighboring GTGGGGA and GATA sites in subunits 2 and 4 to assemble a modular enhancer complex: (E1/NF-Y)–(E2/NF-Y/GATA-2)–(E3/NF-Y)–(E4/MZF1/GATA-2) in K562 cells. In PA-1 cells expressing no GATA-2, the modular enhancer complex appeared to be (E1/NF-Y)–(E2/NF-Y)(E3/NF-Y)–(E4/MZF1/GATAx).

NF-Y bound at the CCAAT motif in the promoter has been reported to be able to open up the chromatin structure of the resident promoters (50, 51). In the trimeric NF-Y complex, the NF-YB and -YC subunits bear both sequence and structure similarities to histone 2B and 2A and can compete with histone 2B and 2A for binding to the histones 3 and 4 tetramer core to disrupt formation of the nucleosomes (59, 60). It is thus possible that in the (4–1)-HS5-ep-GFP plasmid, the e-globin promoter, by itself unable to assemble the NF-Y complex, required the NF-Y delivered from the LTR enhancer to disrupt the nucleosomes and open up its chromatin structure. Indeed, in the integrated (4–1)-HS5-ep-GFP plasmid, the e-globin promoter existed in an accessible chromatin structure and was able to activate transcription of the linked GFP gene at a high level. The ability of the ERV-9 LTR enhancer to transmit its enhancer activity over the intervening HS5 site did not appear to be dependent on the e-globin promoter or special structural features of the (4–1)-HS5-ep-GFP plasmid, because the LTR enhancer coupled to the LTR promoter was able also to act across the HS5 site and other intervening DNAs to activate transcription of the further downstream gene in several different plasmids (15, 16, 37). In contrast, in MI-HS5-ep-GFP, the mutated CCAAT motif in the LTR enhancer could not bind NF-Y to assemble an active LTR enhancer complex. In the absence of NF-Y delivered from the LTR enhancer, the e-globin promoter existed in a closed chromatin structure and was unable to activate transcription of the GFP gene. The results indicate that NF-Y assembled a robust LTR enhancer complex, which could act over the intervening HS5 site to remodel the chromatin structure of the downstream e-globin promoter.

In the endogenous β-globin gene locus, the ERV-9 LTR enhancer spanning a total of 10 NF-Y-binding sites is located near the HS5 site in the 5′ boundary of the β-LCR. Whether the ERV-9 LTR enhancer complex containing multiple NF-Ys can act across the β-LCR to remodel the
chromatin structure of the $\epsilon$-globin promoter located 25 kb downstream of it at the moment not known. However, it has been reported that in the mouse erythroleukemia cells containing a human chromosome 11 spanning the beta-globin gene locus, deletion of hypersensitive sites HS5, -4, -3, and -2 in the beta-LCR, thus juxtaposing the ERV-9 LTR to the HS1 site, shut down transcription of the beta-globin gene but did not close down the chromatin structure of the beta-globin gene locus (61). This finding suggests the possibility that the multiple NF-Ys assembled by the ERV-9 LTR enhancer are able to act over a long distance to remodel the chromatin structure of the beta-globin gene locus. We are testing this possibility by experiments designed to study the effects of ERV-9 LTR deletion on chromatin structure and transcription of the downstream beta-LCR and the globin genes in the 100-kb human beta-globin gene locus during hematopoietic cell differentiation.

Acknowledgments—We thank Dr. A. Phillips for helpful discussion of EMSA results and Dr. R. Markowitz for critical reading of the manuscript.

REFERENCES
1. Henikoff, S., Greene, E., Pietrokovski, S., Bork, P., Attwood, T., and Hood, L. (1997) Science 278, 609–614
2. Lander, E., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., et al. (2001) Nature Genet. 278, 2656–2672
3. Doolittle, W., and Sapienza, C. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 77–83
4. Britten, R. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 9374–9377
5. Ting, C., Rosenberg, M., Snow, C., Samuelson, L., and Meisler, M. (1992) J. Biol. Chem. 267, 14579–14584
6. Schulte, M., Lai, S., Kurtz, A., Czubayko, F., Riegel, T., and Wellstein, A. (1996) J. Biol. Chem. 271, 27600–27605
7. Merika, M., and Orkin, S. (1993) Mol. Cell. Biol. 13, 5454–5463
8. Motta, M., Caretti, G., Badaracco, F., and Mantovani, R. (1999) J. Biol. Chem. 274, 1896–1903
9. Medstrand, P., Landry, J., and Mager, D. (2001) J. Biol. Chem. 276, 31998–32006
10. Henthorn, P., Mager, D., Huisman, D., and Smithies, O. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 5194–5198
11. Lower, R., Lower, J., and Kurth, R. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 5177–5184
12. Temin, H. (1981) Cell 27, 1–3
13. Di Cristofano, A., Strazzullo, M., Parisi, T., and LaMantia, G. (1995) Gene Expression and Hematopoietic Differentiation (Stamatoyannopoulos, G., and Nienhuis, A., eds) pp. 305–322, Alan R. Liss, Inc., New York
14. Zeuken, J., Norgaard, J., Avner, P., Fellous, M., Wartiovaara, J., Valeri, A., Rosen, A., and Giovanella, B. (1980) Int. J. Cancer 25, 19–32
15. Long, Q., Bengra, C., Li, C., Kutlar, F., and Tuan, D. (2002) J. Virol. 76, 1336–1345
16. Medstrand, P., Landry, J., and Mager, D. (2001) J. Biol. Chem. 276, 1896–1903
17. Hromas, R., Collins, S., Hickstein, D., Raskind, W., Deaven, L., O’Hara, P., Magen, S., and Kaufschnsky, K. (1991) J. Biol. Chem. 266, 14183–14187
18. Mitra, J., Hromas, R., and Rauscher, F. (1994) Mol. Cell. Biol. 14, 1786–1795
19. Peteron, M., and Morris, J. (2000) Gene (Amst.) 254, 105–118
20. Tenen, D., Davis, B., Rauchscher, F., Klemsz, A., Tenen, D., Hoffman, S., Xu, D., and Morris, J. (1996) Curr. Top. Microbiol. Immunol. 211, 159–164
21. Tuan, D., Solomon, W., Li, Q., and London, I. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 6384–6388
22. Bieker, J. (2001) J. Biol. Chem. 276, 34355–34358
23. Sambrook, J., and Russell, D. (2001) Molecular Cloning A Laboratory Manual, Vol. 3, 3rd Ed., pp. 1778–1780, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
The Long Terminal Repeat (LTR) of ERV-9 Human Endogenous Retrovirus Binds to NF-Y in the Assembly of an Active LTR Enhancer Complex

NF-Y/MZF1/GATA-2

Xiuping Yu, Xingguo Zhu, Wenhu Pi, Jianhua Ling, Lan Ko, Yoshihiko Takeda and Dorothy Tuan

J. Biol. Chem. 2005, 280:35184-35194.
doi: 10.1074/jbc.M508138200 originally published online August 16, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M508138200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2006/08/15/M508138200.DC1

This article cites 60 references, 35 of which can be accessed free at
http://www.jbc.org/content/280/42/35184.full.html#ref-list-1