THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 11, Issue of March 12, pp. 10523–10531, 2004

Printed in U.S.A.

Deletion of a Negatively Acting Sequence in a Chimeric GATA-1 Enhancer-Long Terminal Repeat Greatly Increases Retrovirally Mediated Erythroid Expression*

Received for publication, December 12, 2003
Published, JBC Papers in Press, December 29, 2003, DOI 10.1074/jbc.M313638200

Anna Testa‡§, Francesco Lotti¶, Linda Cairns‡, Alexis Grande‖, Sergio Ottolenghi‡, Giuliana Ferrari¶, and Antonella Ronchi‡**

From the ‡Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, 20126 Milano, ¶H.S.Raffaele-Telethon Institute for Gene Therapy (HSR-TIGET), 20132 Milano, and the ‖Dipartimento di Scienze Biomediche, Università di Modena e Reggio Emilia, 41100 Modena, Italy

The locus control region of the β-globin gene cluster has been used previously to direct erythroid expression of globin genes from retroviral vectors for the purpose of gene therapy. Short erythroid regulatory elements represent a potentially valuable alternative to the locus control region. Among them, the GATA-1 enhancer HS2 was used to replace the retroviral enhancer within the 3′-long terminal repeat (LTR) of the retroviral vector SFCM, converting it into an erythroid-specific regulatory element. In this work, we have functionally studied an additional GATA-1 enhancer, HS1. HS1 participates in the transcriptional autoregulation of GATA-1 through an essential GATA-binding site that is footprinted *in vivo*. In this work we identified within HS1 a new *in vivo* footprinted region, and we showed that this sequence indeed binds a nuclear protein *in vitro*. Addition of HS1 to HS2 within the LTR of SFCM significantly improves the expression of a reporter gene. The deletion of the newly identified footprinted sequence in the retroviral construct further increases expression up to a level almost equal to that of the wild type retroviral LTR, without loss of erythroid specificity, suggesting that this sequence may act as a negative regulatory element. An improved vector backbone, MΔN, allows even better expression from the new GATA cassette. These results suggest that substantial improvement of overall expression can be achieved by the combination of multiple changes in both regulatory elements and vectors.

GATA-1 is a zinc finger transcription factor that is expressed in a subset of early multipotent and lineage-committed hematopoietic progenitors, including erythroblasts, megakaryocytes, basophils, etc. (1, 2). GATA-1 regulates many transcription factors and lineage-specific genes, and its normal expression is essential for the correct development of several hematopoietic lineages, in particular the erythroid, megakaryocytic, and eosinophilic lineages (3–7).

The regulation of the expression of GATA-1 itself has been the subject of many investigations (7–16). Constructs including the mouse *GATA-1* promoter up to a DNase I-hypersensitive site lying at about −700 nts[1] (HS2) are expressed, at low efficiency, in adult hematopoietic cells in transgenic mice but not in yolk sac cells; however, constructs including an additional more upstream site (HS1) are much more efficient and are also active in primitive hematopoietic cells (12–16). GATA-binding sequences in HS1 and HS2 are essential for activity in a variety of constructs, suggesting GATA-1 auto-regulation (8–10, 14–16); however, a *GATA-1* transgenic construct is active in mice lacking the endogenous *GATA-1* gene, suggesting that other members of the GATA family of transcription factors (possibly GATA-2) may control GATA-1 transcription (13, 16). As GATA-2 is expressed in a wider range of cell types than GATA-1, this implies that other regulatory elements are necessary for appropriate GATA-1 regulation. Sequences relevant for this additional level of regulation have not yet been fully characterized.

Regulatory elements of erythroid genes have been employed in retroviral vectors to express, in model systems, globin genes for therapeutic purposes. In particular, the locus control region (LCR) of the human β-globin cluster, which confers high level, position-independent, erythroid expression of globin genes, gave encouraging initial results (17, 18).

More recently, short functional erythroid enhancers, such as the α-globin HS40 enhancer, the 5-aminolevulinate synthase intron 8, and the GATA-1 HS2 enhancer have been used instead of the β-globin LCR, in various combinations with different erythroid promoters (β and ζ globin, spectrin and ankyrin) to drive gene expression in erythroid cells (19–22). These short elements might thus provide a valuable alternative to LCR for the development of new vectors (23). In this context, the characterization of additional control elements conferring high levels of specific expression is essential for vector optimization. Among the enhancers studied, the GATA-1 HS2 element showed significant activity both in retroviral and lentiviral vectors (19–21).

In this work we have further characterized functional sequences of the GATA-1 HS1 enhancer. By adding HS1 to HS2

* This work was supported by a core grant from the Italian Telethon Foundation (to G. F.) and Cofin 2002 (to S. O. and F. Mavilio, head of the group at the Università de Modena e Reggio Emilia). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Present address: Dipartimento di Scienze Biomediche, Università di Modena e Reggio Emilia, 41100 Modena, Italy.

** To whom correspondence should be addressed: Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, Italy. Tel.: 3902-6448-3337; Fax. 3902-6448-3314; E-mail: antonella.ronchi@unimib.it.

[1] The abbreviations used are: nts, nucleotides; LTR, long terminal repeat; LCR, locus control region; DMS, dimethyl sulfate; EMSA, electrophoretic mobility shift assays; BM, bone marrow; Epo, erythropoietin; FCS, fetal calf serum; FACS, fluorescence-activated cell sorting; MEL, mouse erythroleukemia; wt, wild type; IVFP, *in vivo* footprinting protein; YS, yolk sac; MFI, mean fluorescence intensity; GFP, green fluorescent protein; EGFP, enhanced GFP; NGFr, nerve growth factor receptor; ts, thermosensitive.

within the retroviral LTR and by deleting an inhibitory sequence within HS1, we obtained retroviruses that express downstream reporter genes at similar efficiencies as those retaining the wild type LTR, without loss of erythroid specificity.

EXPERIMENTAL PROCEDURES

In Vivo Footprinting—*In vivo* DMS treatment of cells, DNA extraction, and piperidine treatment were according to Ref. 24; ligation-mediated PCR was according to Ref. 25. Primers used are as follows: P1 from nt 2 to 25 (GATCCAAGGAAGAGAGGACATTAG); P2 from nt 13 to 36 (AGAGGACATTAGCATGGGTCTCAA); and P3 from nt 27 to 50 (GGGTCTCAAATGGAAGCCTGACAG) of HS1 (14) in the sense orientation.

Electrophoretic Mobility Shift Assay (EMSA)—Nuclear extracts were prepared as in Refs. 26 and 27; *in vitro* binding and gel electrophoresis were as in Refs. 27 and 28.

The sequences of the oligonucleotides used are as follows: +46+96 wt, 5′-CTGACAGAGAAGA**CGCTTCAACCC**GGACACCCCACCCCC-GCCTGCAATGGG-3′; Δ59–69, 5′-CTGACAGAGAAGAGGACACCCC-ACCCCCGCCTGCAATGGG-3′; core mut, 5′-CTGACAGAGAAGACGC-TTTAATTTGGACACCCCACCCCCGCCTGCAATGGG-3′; 5′mut, 5′-C-TGACAGTTAATTTGCTTCAACCCGGACACCCCACCCCCGCCTGCA-ATGGG-3′; +81+111, 5′-CCCGCCTGCAATGGGCTCCCCCAAGCCT-AG-3′; and β-globin CACC box, 5′-CTTGGGGGCCCCTCCCCCACAC-TATCTCAA-3′.

The sequence in boldface type in the wild type oligonucleotide corresponds to the *in vivo* footprinted region (see "Results") and is deleted in the Δ59–69oligo. Underlined nucleotides represent the mutated sites.

Retroviral Constructs—The retroviral vectors SFCM, containing an intact Moloney enhancer and its derivative Δ·SFCM (in which the 3′-LTR enhancer is deleted), have been described previously (19).

The HS2 site of the mouse *GATA-1* gene was inserted in place of the deleted enhancer into the Δ·SFCM vector to generate GATA·SFCM (19), here renamed as HS2·SFCM. Additional constructs were generated by inserting appropriate HS1 fragments into either Δ·SFCM or HS2·SFCM. The genomic 461-nt BamHI BglII HS1 fragment (14) was modified with XhoI linkers and cloned into the compatible SalI site in the Δ·SFCM LTR polylinker, 10 nt upstream to the SnaBI site where HS2 was originally cloned. In the constructs carrying both HS1 and HS2, HS1 was inserted in the same SalI site into the HS2·SFCM vector to generate HS1·HS2·SFCM.

All HS1 mutants (see the above oligonucleotides for mutations) were generated by PCR and sequenced to confirm the mutation. In the HS1GATA$^-$·HS2·SFCM vector the HS1 GATA-1-binding site at positions 130–136 (14) was changed from CTTATC to CTTAAA.

The LGSΔN vector was derived by replacing the TKNeoR sequence in the SFCMM-3 vector (29) with the 730-bp Eco47III-ScaI EGFP fragment from pEGFP-C3 (Clontech). HS2·LGSΔN was obtained by cloning the BamHI 200-bp GATA-1 HS2 cassette (19) into the SalI/BglII sites of LGSΔN, whereas ΔHS1·HS2·LGSΔN was derived by ClaI/NdeI digestion of the ΔHS1·HS2 elements from the ΔHS1·HS2·SFCM vector and cloning into ClaI/NdeI sites of LGSΔN.

Cells—GATA-1 ts Epo bone marrow (BM) and GATA-1 ts Epo YS (yolk sac) are erythropoietin (Epo)-dependent immortalized hematopoietic progenitors derived from bone marrow and yolk sac of mice transgenic for a mutant SV40 T gene encoding a thermosensitive T protein driven by the *GATA-1* promoter linked to either the HS2 or the HS1 sites, respectively (11, 30).[2] These cells grow at 32 °C, the permissive temperature for the thermosensitive T protein with a doubling time of 20–24 h, in RPMI 1640, 10% fetal calf serum (FCS), 0.3 unit/ml Epo.

Transduction of Mouse and Human Cell Lines—For transduction of mouse cells, packaging lines for SFCM vectors were obtained by plasmid transfection of the ecotropic line GP$^+$E86 (31) and selection in 0.8 mg/ml G418. For viral infections, subconfluent packaging lines were treated with mitomycin C (1 μg/ml) for 3 h at 37 °C, carefully washed, and replated (1×10^6 cells) into 6-cm diameter dishes in 3 ml of RPMI 1640, 10% FCS. The next day, $0.6–1 \times 10^6$ mouse hematopoietic cells in 1 ml of complete Epo-containing medium were added to each dish and grown for 24 h at 32 °C. The hematopoietic cells growing in suspension were collected, washed, and grown for 2 days prior to selection. G418 was then added at 0.8 mg/ml, and resistant cells were obtained within 10–15 days.

Selected cells were then analyzed by flow cytometry (see below). For each infection, the copy number of constructs integrated within the resistant cells was evaluated by Southern blotting by using a probe for the neomycin resistance gene and checked for equal loading by further probing with a genomic fragment downstream of the GATA-1 locus. The copy numbers were essentially identical between all samples and close to one integration (on average) per cell (as determined by comparison with serial dilutions of appropriately digested viral DNA mixed with genomic DNA of non-infected cells).

NIH3T3 cells were grown in Dulbecco's medium (Invitrogen) supplemented with 10% FCS (Hyclone) and infected for 16 h with undiluted viral supernatants containing 8 μg/ml Polybrene (Sigma). Further analyses were carried out on transduced and control bulk cultures.

For transduction of human cells viral stocks were produced by transient transfection of Phoenix-Ampho cells as described previously (32). Human erythroblastic K562 and HEL cell lines (ATCC) were grown in RPMI 1640 (Invitrogen) supplemented with 10% FCS and transduced by spinoculation (33) in the presence of Polybrene (8 μg/ml). For transduction of mouse hematopoietic cells with LSGΔN and MΔN viral stocks, spinoculation was used as above.

FACS Analysis—Expression of the reporter gene (ΔLNGFr) was monitored by flow cytometry (FACScan, BD Biosciences) using the murine anti-human p75-NGFr monoclonal antibody 20-4 (ATCC). A goat monoclonal anti-mouse IgG (Fab-specific) conjugated to fluorescein isothiocyanate or phycoerythrin was used as secondary antibody (Pharmingen).

RESULTS

An "In Vivo" Footprint in GATA HS1—DNA fragments from the HS1 region of the mouse *GATA-1* gene have been shown previously to have enhancer activity, when linked to the *GATA-1* promoter alone or in combination with HS2 and other GATA-1 regulatory elements, in transient transfections, and in transgenic assays (13–16). The activity of HS1 is totally dependent on a strong GATA-1-binding site (14–16) that appears to be occupied *in vivo*, as indicated by *in vivo* footprinting (14). To identify additional functional motifs within HS1, we further extended the dimethyl sulfate (DMS) *in vivo* footprinting analysis of HS1 (Fig. 1*A*).

Transcription factors bound to specific DNA sequences protect guanines from *in vivo* DMS-induced methylation and thus from subsequent *in vitro* piperidine cleavage of the methylated sites. The following ligation-mediated PCR results in the absence of bands at the positions corresponding to the guanines protected "*in vivo*" (24, 25).

Fig. 1*B* shows that six guanines (lower strand) within an 11-nt sequence between positions 59 and 69 of HS1 are footprinted "*in vivo*" in hematopoietic mouse erythroleukemia cells (MEL) grown in culture; a weaker footprint is also detected on the CACCC motif between positions 78 and 82 (not shown).

In Vitro Binding Studies of the Footprinted Region—To better characterize proteins responsible for the *in vivo* footprint, we performed EMSA. A ^{32}P-labeled oligonucleotide (wt, see "Experimental Procedures") spanning from nt 46 to 96 generates a complex pattern consisting of an intense slow band and additional faster bands in the presence of nuclear extracts from the hematopoietic cell lines: human K562 erythroleukemia cells, murine GATA-1 ts Epo BM, and GATA-1 ts Epo YS cells (30), (Fig. 2); however, when an oligonucleotide deleted between nucleotides 59 and 69 (Δ59−69)is used, one of these bands is clearly missing (Fig. 2*B, arrow*, compare *lanes 1* and *2*). The protein responsible for this band will be called from now on the GATA-1 HS1 "*in vivo* footprinting protein" (IVFP). As expected, unlabeled oligonucleotides including a CACCC box sequence from the human β-globin gene and the GC-rich region (positions 81–111) from the HS1 itself compete all bands, with the exception of the IVFP band (Fig. 2*B, lanes 3* and *4* and *lanes 7* and *8*); an unrelated control oligonucleotide has no effect (*lanes 5* and *6*).

We further tested two additional oligonucleotides (Fig. 2): "core mutant" carries mutations centered on nucleotides involved in the *in vivo* footprint, whereas the "5′ mutant" is

[2] L. Cairns, S. Ottolenghi, and A. Ronchi, unpublished data.

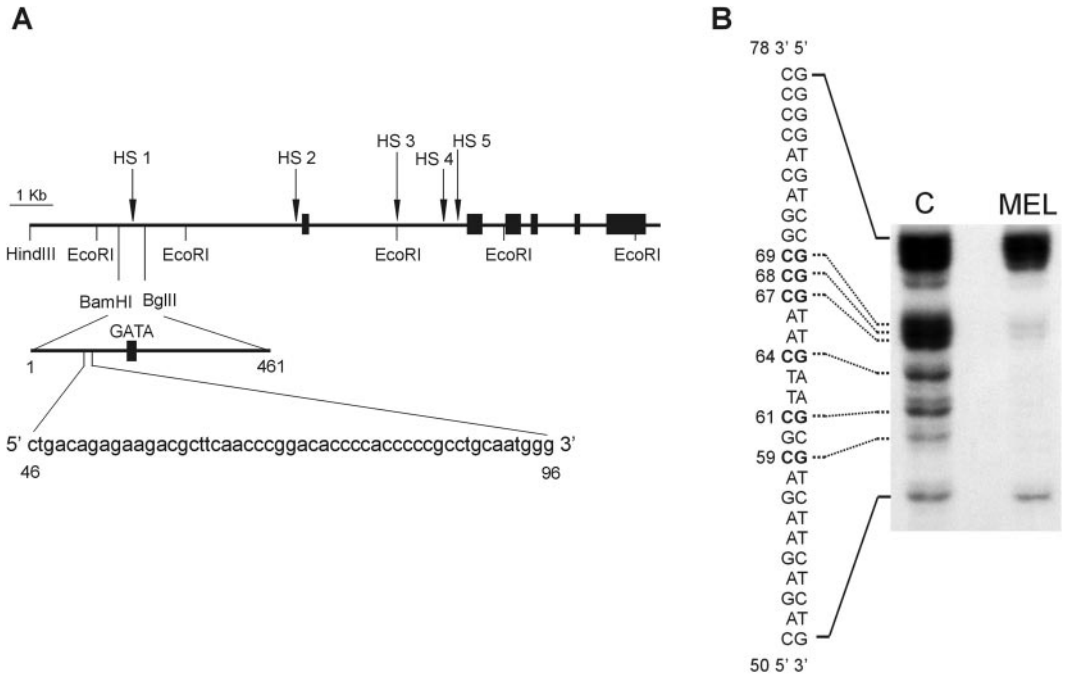
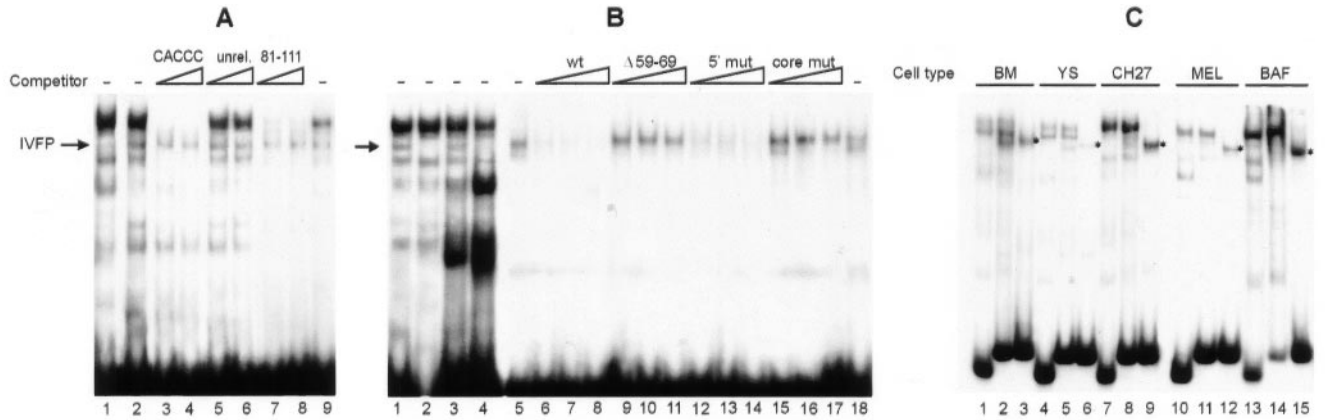


FIG. 1. **An *in vivo footprinted site within the HS1 region of the mouse GATA-1* gene.** *A*, schematic representation of the *GATA-1* gene. The DNA sequence (*top strand*) encompassing the *in vivo* footprint is expanded; *numbers* refer to nucleotide positions in Ref. 14. *B*, the footprinted region. Guanines protected from methylation in the erythroid MEL cell line are shown in *boldface* and correspond to the *bottom strand* in *A*. *Lane C* is from DMS-treated naked DNA.



+46+96 wt	5′ CTGACAGAGAAGA**CGCTTCAACCC**GGACACCCCACCCCCGCCTGCAATGGG3′
Δ59-69	5′ CTGACAGAGAAGAGGACACCCCACCCCCGCCTGCAATGGG3′
core mut.	5′ CTGACAGAGAAGACGCTTTAATTTGGACACCCCACCCCCGCCTGCAATGGG3′
5'mut.	5′ CTGACAGTTAATTTGCTTCAACCCGGACACCCCACCCCCGCCTGCAATGGG3′
+81+111	5′ CCCGCCTGCAATGGGCTCCCCCAAGCCTAG3′

FIG. 2. **EMSA analysis of the footprinted region.** *A*, the deletion of nucleotides 59–69 within a ^{32}P probe encompassing nt 46–96 abolishes a single band (IVFP). *Lane 1*, deleted probe. *Lanes 2–9*, wt probe. Unlabeled competitor oligonucleotides (50- and 150-fold excess) were added (*top*), as indicated. *CACC* is a human β-globin CACC box oligonucleotide; *81–111* oligonucleotide includes the corresponding sequences of the HS1 region, and *unrel* is a nonspecific oligonucleotide carrying an Oct-1-binding site. *B*, a mutation within the 59–69 nucleotide region abolishes the IVFP band. *Lane 1*, wt probe; *lane 2*, Δ59–69 probe; *lane 3*, 5′ mutation; *lane 4*, core mutation. *Lanes 5–18*, binding of the wt probe in the presence of poly(dG·dC) instead of poly(dI·dC). Competitors (50-, 150-, and 300-fold molar excess) are indicated on the *top* of the figure. Nuclear extracts were from erythroid K562 cells. *C*, the IVFP protein is present in a variety of mouse hematopoietic cells. Each extract was tested with the Δ59–69 probe and with the wt 46–96 probe in the presence of poly(dI·dC) and with the same wt probe in the presence of poly(dG·dC) instead of poly(dI·dC) (*lane 1–3*, respectively, for each nuclear extract). The *asterisk* indicates the IVFP band.

mutated at the upstream border of the footprinted region (see "Experimental Procedures" for sequences). The mutation in oligonucleotide core mutant (Fig. 2*B, lane 4*) has the same effect as the deletion of nucleotides 59–69, *i.e.* loss of IVFP binding, whereas the upstream mutation in oligonucleotide 5′ mutant does not significantly affect IVFP binding (Fig. 2*B, lane 3*).

To better resolve the IVFP band, we used in EMSA experiments poly(dC·dG) as a competitor, instead of poly(dI·dC) (Fig. 2*B, lanes 5–18*). Under these conditions, only the IVFP band is visible (*lane 5*) with the wt oligonucleotide and is efficiently competed by the unlabeled wt oligonucleotide itself (Fig. 2*B, lanes 6–8*) and by the 5′ mutant oligonucleotide (*lanes 12–14*)

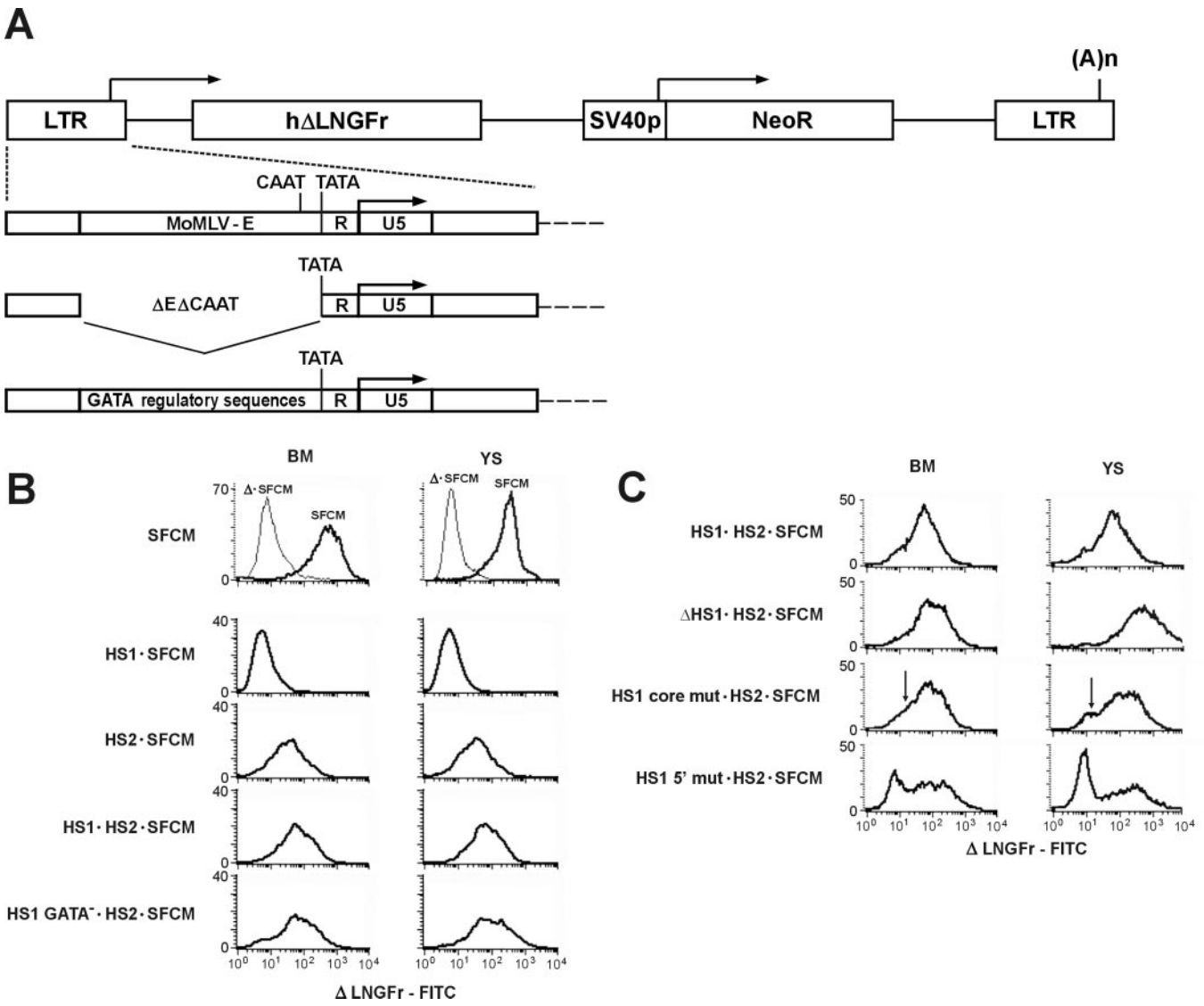


FIG. 3. **FACS analysis of LTR-dependent Δ*LNGFr* reporter gene expression from SFCM retroviral vectors containing HS1 sequences.** *A*, schematic representation of the SFCM retroviral vector. GATA-1 regulatory sequences were inserted in a retroviral enhancer-deleted version of the vector (see expansion). *B* and *C*, Δ*LNGFr* gene expression in erythroid cells transduced with modified vectors (see text). *BM* and *YS* indicate the immortalized multipotent cell lines used. *B*, *SFCM panel*, *thin* and *thick lines* indicate Δ·SFCM- and SFCM- transduced cells, respectively. In control experiments (not shown), the profile of untransduced cells is essentially identical to that of Δ·SFCM-transduced cells. Similarly, the use of the second antibody alone gives no significant staining.

but not by the deleted (*lanes 9–11*) or core mutant (*lanes 15–17*) oligonucleotides.

The IVFP protein appears to be relatively widespread, although its amount varies between different cell types. In particular, it is expressed in both mouse and human cells, particularly in human K562 and mouse erythroleukemia MEL cells, in SV40-T immortalized bone marrow (GATA-1 ts Epo, BM) and yolk sac (GATA-1 ts Epo, YS) erythropoietin-dependent cells (11, 30),[2] and in lymphoid cells (CH27, A20, and BAF3) (Fig. 2*C* and data not shown). By using mutant oligonucleotides with these extracts, we obtained results similar to those shown for K562 (data not shown).

HS1 DNA Sequences Co-operate with HS2 in Retrovirally Mediated Gene Expression in Hematopoietic Cells—To analyze functional effects of the HS1 DNA sequences, we used a modified SFCM retroviral vector, as described previously (19). In this vector, the Moloney-leukemia virus enhancer in the 3′-LTR is deleted (Δ·SFCM) and can be replaced by an exogenous enhancer; following retroviral infection of the target cells, the new enhancer is moved to the 5′-LTR from which it drives retroviral transcription and expression of a truncated form of nerve growth factor receptor (ΔLNGFr), a reporter gene. In addition, an internal SV40 promoter drives the expression of a neomycin resistance gene (Fig. 3*A*). Our strategy was therefore to infect *in vitro* hematopoietic cell lines with appropriate constructs and to select cells expressing the neo-resistance gene by G418 treatment. Following complete selection, the proportion of expressing cells and levels of expression were analyzed by fluorescence-activated cell sorting (FACS).

The copy number of integrated constructs was examined by Southern blotting using a neo-resistance probe (for the detection of the retroviral DNA) and a GATA-1 3′ genomic probe for detection of an endogenous gene and normalization. The copy numbers were very similar between the various constructs and close to 1 copy per cell, as expected for the relatively low probability of infection in these experiments (between 25 and 50% of cells transduced).

For these experiments, we primarily used two cell lines, GATA-1 ts Epo (BM) and GATA-1 ts Epo (YS) (11, 30),[2] which represent erythropoietin-dependent multipotent cells derived

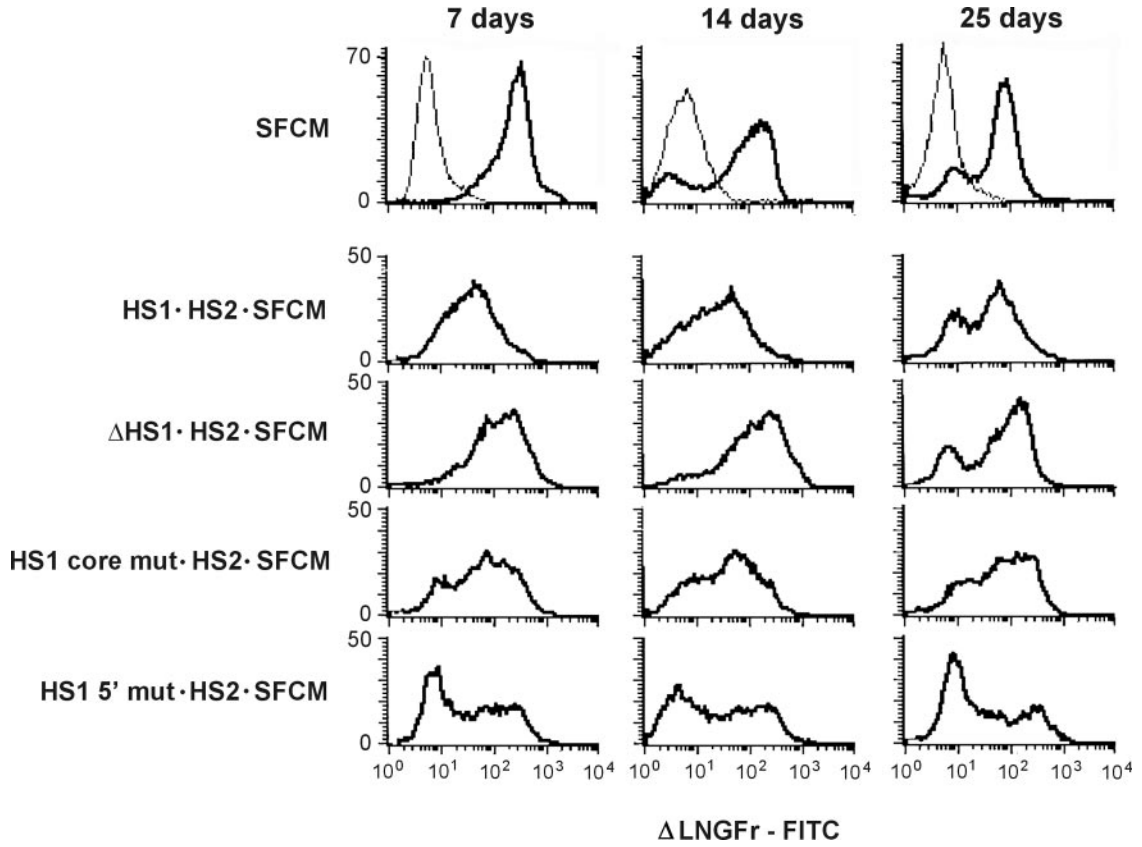


FIG. 4. **FACS analysis of transduced and G418-selected cells after withdrawal of G418.** Yolk sac cells retrovirally transduced with the indicated constructs were selected in G418 for more than 3 weeks and then analyzed at various times after withdrawal of G418.

from bone marrow and yolk sac cells immortalized *in vitro* by SV40 T-antigen (driven by a transgenic *GATA-1* promoter linked to HS2 or HS1 respectively). These cells are 95–100% positive for endogenous GATA-1 (based on nuclear immunofluorescence tests) (30) and are thus expected to express foreign constructs depending on GATA regulatory elements.

In a first series of experiments (Fig. 3*B*), we compared the activities of retroviral constructs (HS1·SFCM and HS2·SFCM) containing HS1 or HS2 as the foreign enhancer replacing the original Moloney enhancer. The mean fluorescence intensity (MFI) of cells infected with the wild type SFCM is given as 100% activity. The MFI of each construct is given below as percentage of activity relative to that of the wild type SFCM. The activity of HS1·SFCM (2 ± 1% in both BM and YS cells) and mutated derivatives of it (data not shown) did not differ from that of the inactive control Δ·SFCM (2 ± 1% of SFCM). On the other hand, HS2·SFCM was clearly active in both BM and YS cell lines (as expected, see Ref. 19) at moderate levels (10 ± 2 and 11 ± 2% of SFCM, respectively).

We then linked HS1 upstream to HS2 (HS1·HS2·SFCM); this resulted in a ≈2-fold increase ($p < 0.05$ by Student's t test) of the activity as compared with HS2·SFCM (19 ± 3% of SFCM in BM and 21 ± 2% in YS), indicating that HS1 sequences synergize with HS2. Interestingly the same result was obtained when HS1 carrying a deletion of its 3′ sequences (nt 1–171, data not shown) or a point mutation in the GATA-1 motif (HS1GATA$^-$·HS2·SFCM, Fig. 3*B*) was used instead of the intact HS1. Similar results were obtained with GATA-1 ts Epo (BM) and (YS) cells (Fig. 3*B*).

Deletion of the IVFP-binding Site Increases the Activity of the HS1·HS2·SFCM Construct—We tested the effects of the deletion of the *in vivo* footprinted elements (Fig. 3*C*). Although a vector carrying HS1 deleted at nucleotides 59–69 (ΔHS1·SFCM) is completely inactive on its own (data not shown), the addition of the HS1Δ59–69 element to HS2 (ΔHS1·HS2·SFCM vector) caused a strong increase of reporter gene expression in both bone marrow and yolk sac cell lines. Most interesting, in the latter, the effect on activity of the addition of HS1 was more pronounced (56 ± 12% *versus* 100 ± 20%, respectively).

To ascertain whether the loss of the IVFP-binding site could contribute to this increase, we inserted into HS1 the same mutations as those tested in EMSA experiments. Introducing the core mutation in HS1·HS2·SFCM, increases the average level of activity of the construct (HS1core mutant·HS2·SFCM, 37 ± 14% of SFCM in BM and 40 ± 21% in YS), although not as much as the deletion. Note that in the core mutant-infected cells, many (80%) of the cells express Δ*LNGFr* at high levels, as in ΔHS1·HS2·SFCM, but a significant shoulder of non expressing cells is present. In contrast, the vector (HS1 5′ mutant·HS2·SFCM) carrying the 5′ mutation that slightly affects IVFP binding (Fig. 2*B*) is active in a much smaller proportion of cells, although some still express at high levels.

Exogenously inserted genes are often silenced during long term cell propagation in culture, as a repressive chromatin structure adjacent to the integration site might extend into the foreign gene (34). When this event occurs, both the reporter gene and the neomycin-resistance gene are likely repressed (34). Under these conditions, the presence of G418 in the medium will cause loss of these cells, leading to an overestimate of the proportion of reporter-expressing cells. We thus tested whether releasing the cells from G418 selection affects reporter gene expression (Fig. 4). Cells grown in G418 for at least 3 weeks to kill all non-virus-infected cells were further propagated in the absence of G418 and tested for reporter gene expression. With all tested constructs, including the wild type Moloney enhancer-dependent SFCM construct, there was some degree of progressive inactivation, as expected (34–36); how-

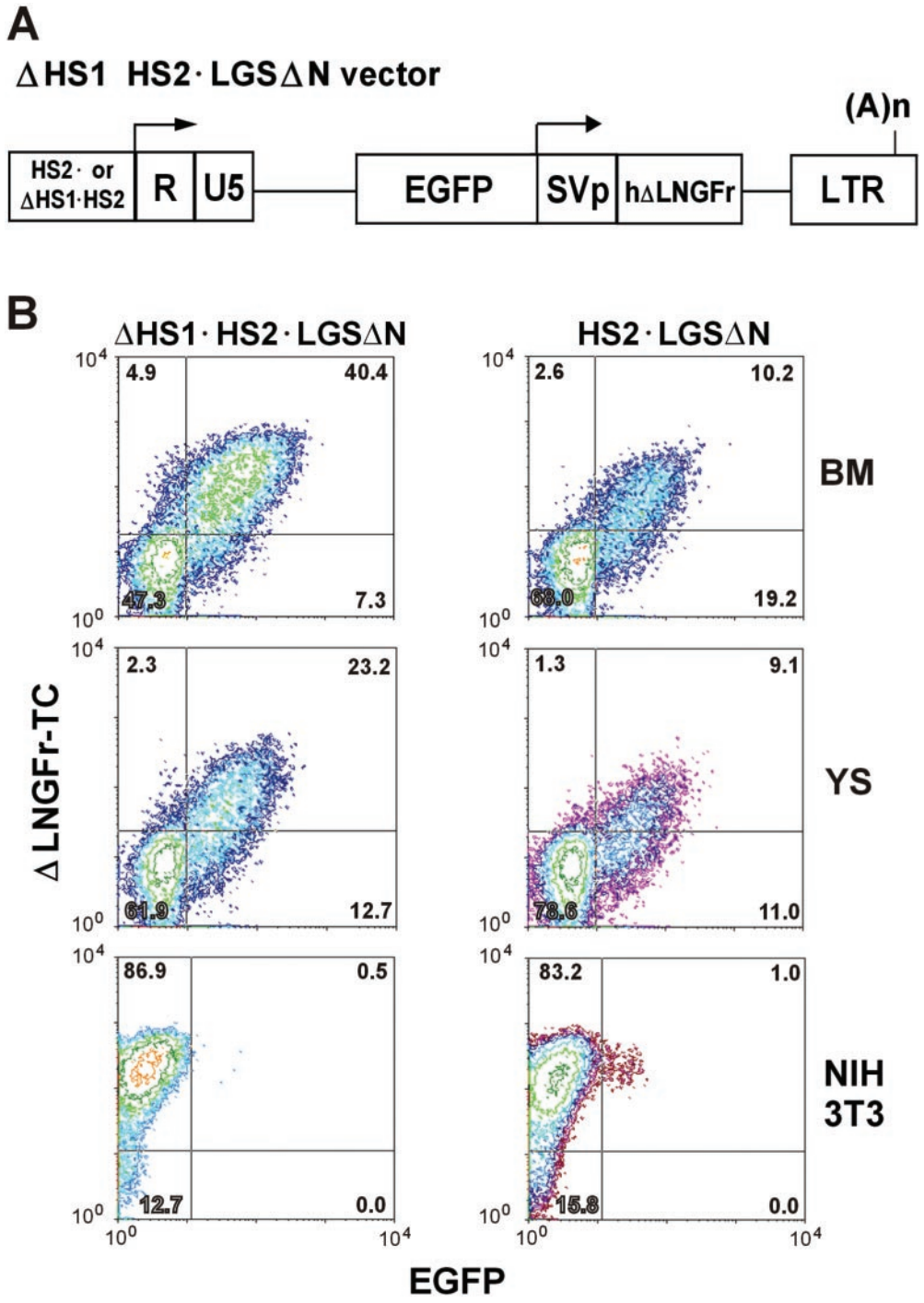


FIG. 5. **Erythroid restriction of GATA-1-modified LTR-dependent reporter EGFP gene expression.** *A*, the GATA-1-modified LGSΔN vector. *SVp* indicates SV40 promoter. *B*, FACS analysis of transduced hematopoietic mouse GATA-1 ts Epo BM, GATA-1 ts Epo YS cells, and mouse fibroblast 3T3 cells.

ever, after 25 days in culture without G418 (~30 cell cycles), the reporter gene was still active at high levels in 75–90% of cells with all constructs, except for HS15′ mutant·HS2·SFCM which remains active in only 25% of the cells (Fig. 4). This demonstrates that SFCM or HS1·HS2·SFCM, ΔHS1·HS2·SFCM, and (to a lesser extent) HS1core mutant·HS2·SFCM-modified LTR are relatively well protected from position of integration effects. The mutation of the 5′ sequences within the HS1·HS2 context, however, clearly interferes with the maintenance of the activity.

HS1 Increases Gene Expression in Erythroid Cells but Not in Fibroblasts—HS2·SFCM has been shown previously to drive reporter gene expression in hematopoietic but not non-hematopoietic cells (19). To ascertain if the ΔHS1·HS2 cassette retains hematopoietic specificity, we constructed an additional vector (ΔHS1·HS2·LGSΔN) in which the modified LTR drives the activity of GFP, whereas a constitutive internal promoter expresses the Δ*LNGFr* gene, acting as an internal reference for the level of ubiquitous expression in uninduced cells (Fig. 5*A*). This experimental arrangement allows us to measure LTR-dependent gene expression in the absence of any previous selection. We thus compared the transgene expression level in cells transduced with the HS2·LGSΔN vector and the ΔHS1·HS2·LGSΔN vector (Fig. 5*B*). Mouse NIH3T3 fibroblasts, mouse hematopoietic GATA-1 ts Epo BM and GATA-1 ts Epo YS cells, and human

erythroid K562 cells were transduced with HS2·LGSΔN and ΔHS1·HS2·LGSΔN viral supernatants, maintained in culture as bulk population, and assayed for ΔLNGFr and GFP expression by FACS analysis (Fig. 5*B*).

Most (over 80%) of the NIH3T3 cells transduced with either HS2·LGSΔN or ΔHS1·HS2·LGSΔN expressed the Δ*LNGFr* gene, at similar levels (Fig. 5*B*). In contrast, very few cells also expressed GFP, at low levels, indicating that the GATA-1 cassettes are essentially inactive in NIH3T3 cells. Hematopoietic GATA-1 ts Epo BM and GATA-1 ts Epo YS cells were transduced with the same viral supernatants as above, to allow comparison. In contrast to NIH3T3 cells, most of the hematopoietic NGFr-positive cells were also strongly GFP-positive; the fact that some cells with very low levels of NGFr expression do express some GFP suggests that the erythroid cassette may be more easily activated within hematopoietic cells than the SV40 promoter driving NGFr. As expected, the proportion of GFP-expressing cells was much higher in ΔHS1·HS2·LGSΔN than in HS2·LGSΔN transduced cells.

In similar experiments, in human erythroleukemic K562 cells transduced with ΔHS1·HS2·LGSΔN, the GFP was expressed at MFI (2.5-fold higher than that in cells transduced with HS2·LGSΔN (not shown). These results confirm that erythroid restriction of transgene expression by the HS2 element of the *GATA-1* gene is maintained also when this element is used in combination with the ΔHS1 element of the *GATA-1* gene.

An Improved Vector Backbone Further Increases Gene Expression Activity—Another approach to increase the overall efficiency of transgene expression was to improve the vector characteristics at the post-transcriptional level. We compared the ΔSFCM backbone (originally derived from the L*X*SN vector) with that of the MΔN vector (originally derived from MFG (37)), in which the primary genomic transcript is spliced from a combination of the viral *gag* donor and *env* acceptor sites, and the expressed cDNA is cloned in-frame with the *env* start codon and translated under the control of the highly efficient *env* leader sequence. The U3 LTR sequence in MΔN vector (38), containing the Δ*LNGFr* as marker gene, was deleted and replaced by the chimeric LTR from the ΔHS1·HS2 ·ΔSFCM vector, generating the ΔHS1·HS2·MΔN vector. The improvement of gene expression obtained with these modifications was investigated at the level of RNA accumulation by Northern blot analysis and of protein synthesis by FACS analysis. Mouse GATA-1 ts Epo BM cells and human erythroid HEL cells were transduced, using the same conditions of infection, with ΔHS1·HS2·SFCM and with ΔHS1·HS2·MΔN viral supernatants and grown in culture as bulk population. Transduction efficiency with the ΔHS1·HS2·SFCM and ΔHS1·HS2·MΔN vectors was above 80% in all cell lines. ΔLNGFr expression, in terms of MFI, was 2–5-fold higher in cells transduced with ΔHS1·HS2·MΔN vector than in those transduced with the HS2·SFCM vector and approached that of the wild type MΔN vector. The increased expression observed with MΔN vector was at the level of RNA accumulation (data not shown).

DISCUSSION

DNA Sequences Regulating GATA-1 Expression—The DNA region surrounding the HS1 site of the mouse *GATA-1* gene appears to play a central role in GATA-1 expression in transgenic experiments, as it is essential for high level activity in hematopoietic cells in general, and in particular for expression in embryonic erythroblasts (12–16). It is likely that a certain degree of functional redundancy exists in the *GATA-1* gene region; in fact, the deletion from the endogenous *GATA-1* gene of a region comprising HS1, by homologous recombination, has little effect on GATA-1 expression in erythroid cells (7). Nevertheless, characterizing functional elements of the GATA-1 HS1 region may lead to the discovery of important regulatory factors for GATA-1 expression. In this regard, it is relevant that the single HS2 site, in the absence of any additional GATA-1 regulatory elements, was able to confer proper expression of a reporter gene within the SFCM retrovirus in stem cell-derived early hematopoietic progenitors and erythroblasts (19).

Cooperation between HS1 and HS2—Our results show that HS1, unexpectedly, is unable to direct hematopoietic cell expression of a reporter gene when inserted alone into the SFCM LTR. This is in contrast with evidence obtained in transient and stable transfection experiments by ourselves and others (14–16). However, the same element significantly stimulates the activity of a linked HS2 sequence. The HS1 element contains a GATA-binding motif that might bind either GATA-1 or GATA-2 and that might play a role in an autoregulatory loop and/or in the initial activation of the *GATA-1* gene itself (14–16). In fact, the HS1 GATA-binding sequence was found to be essential in transient transfection assays of HS1 alone (14) as well as in transgenic constructs carrying HS2 and other regulatory elements (12, 13, 15, 16) within wild type or even GATA-1$^-$ mutant mice. In our experiments, HS1 synergizes with HS2 (Fig. 5) even when the HS1 GATA-binding site is mutated; this suggests that sequences other than the sole GATA-binding site are important within HS1 for its functional cooperation with HS2. The discrepancy between the previously described requirement for the GATA-binding site and our present data might be explained by the different arrangement of HS1 and HS2 in the transgenic *versus* the retroviral constructs; in the latter one the HS1 and HS2 sites are immediately adjacent, and the strong GATA-1-binding element of HS2 might be sufficient for activity, whereas in the transgenic constructs HS1 and HS2 are separated by almost 3 kb and may thus act independently.

Negatively Acting Elements—In vivo footprinting studies have so far identified within HS1 only two sites that are footprinted *in vivo*, thus directly implying a significant functional role: the GATA-binding site (14) and the upstream 59–69 footprint (see above). The latter footprint is immediately adjacent to a GC-rich region that binds SP1 *in vitro* and the erythroid EKLF factor weakly (data not shown) and yields a weaker footprint. Binding sites for other (uncharacterized) factors also lie upstream to the footprint. When HS1 deleted in the footprinted region is added to HS2, the resulting construct ΔHS1·HS2·SFCM is more active than the undeleted HS1·HS2·SFCM.

It is interesting that ΔHS1·HS2 expression is stronger in the yolk sac-derived cell line than in the bone marrow line immortalized by a T gene dependent on HS2 activity. This is in keeping with the notion that HS1 is strictly required for GATA-1 expression in yolk sac cells (12, 13, 15, 16).

The functional effect of the deletion of the 59–69 region (Fig. 3) can be reproduced, at least in part, by a mutation (core mutant) that destroys the binding of IVFP (Fig. 2*B*). Thus a specific ubiquitous protein, IVFP, might negatively regulate HS1 activity (or its ability to cooperate with HS2) in our experimental system. However, it is possible that the stronger effect of the deletion (*versus* the core mutation) might also be due to a perturbation of the architecture of the region introduced by the deletion itself. The deletion might favor the functional cooperation of factors binding to the sequences flanking the footprinted region by joining them into a single DNA stretch (Fig. 1*A*). It is of further interest that the HS1 core mutation, while improving reporter expression in a large subset of cells, also significantly increases the number of cells that fail to express (Fig. 3*B*).

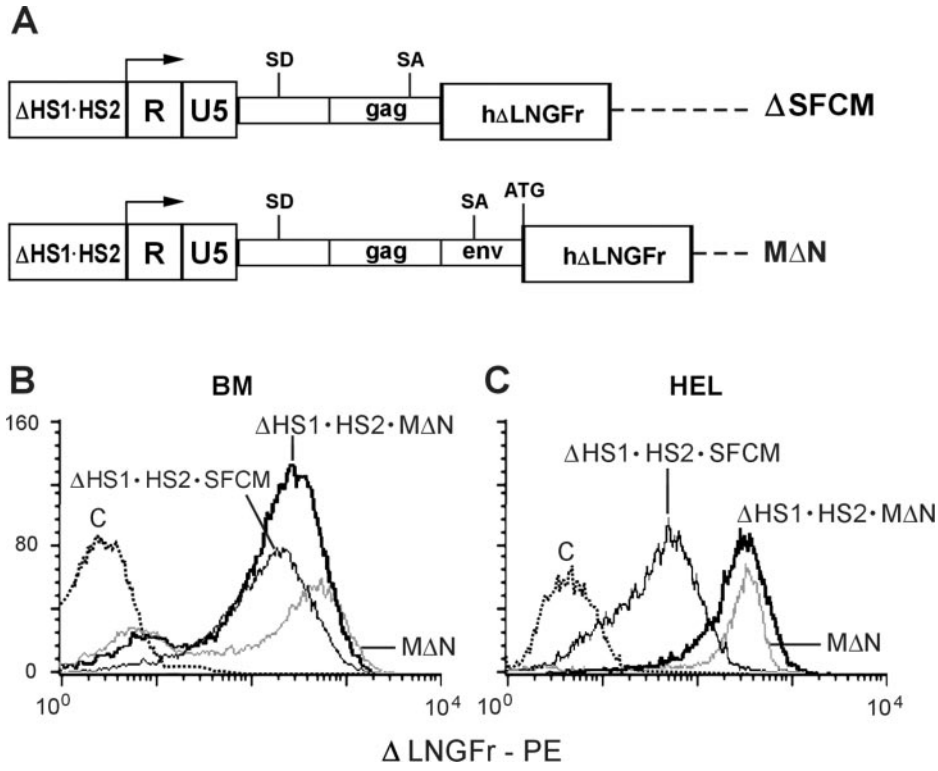


FIG. 6. **Improved ΔLNGFr expression in the MΔN-based vector.** *A*, comparison of the SFCM and MΔN backbones. *B*, FACS analysis of transduced mouse GATA-1 ts Epo BM and human erythroid HEL cells with different vectors. Untransduced BM and HEL cells are indicated by "*c*" in *B* and *C*, respectively.

Similarly, the HS1 5′ mutation results in a bimodal distribution of reporter activities, with a large proportion of nonexpressing cells (Fig. 3*B*). Intriguingly, HS2·SFCM, although active at relatively low levels, is expressed in the majority of the cells (Fig. 3*A*). These results indicate that sequences within HS1, when mutated, might interfere with the activity of the adjacent HS2 element. It is possible that the IVFP binding region and flanking sequences include elements affecting "chromatin opening" or boundaries shielding HS1 from the effects of neighboring sequences. Alternatively, the mutations introduced (but not the deletion) might have inadvertently generated negatively acting elements. Future methylation studies of integrated sequences carrying normal and mutated IVFP binding region might help to solve this problem.

We wish to stress that because of the differences between the relative arrangement of HS1 and HS2 sequences within the retroviral constructs and their *in vivo* location, we cannot infer from these data the *in vivo* role of the IVFP-binding sequence in GATA-1 regulation; however, the *in vivo* footprint points to a significant functional role within hematopoietic cells. In future work we will try to identify additional binding sites for IVFP within the mouse *GATA-1* gene, and possibly other erythroid genes, and to assess the functional role of the footprinted region by transgenic experiments. From the evolutionary point of view, the IVFP-binding site is very weak or lost in the human HS1, but other possible binding sites within the human *GATA-1* gene have not been studied yet. Interestingly, a perfect (17/17 nt) match between the footprinted region in GATA-1 and a mouse activin receptor type II sequence immediately downstream to the promoter was detected; this sequence efficiently binds IVFP *in vitro*, thus suggesting a possible role for IVFP in the regulation of this gene (not shown).

Improvement of Erythroid-specific Retroviral Vectors—The traditional approach to gene therapy of inherited hemoglobin disorders has relied on the use of globin gene LCR-derived cassettes, recently achieving promising results (17, 18). The large size of the fragments necessary for optimal LCR activity (*i.e.* expression level and full insulation from position effects) has been an obstacle to the formation of high titer, high efficiency vectors which do not undergo structural rearrangements. As an alternative approach, other erythroid regulatory elements have been used instead of the LCR (19–23), including the GATA-1 HS2 enhancer. Its relatively small size allowed us to insert it directly into the LTR, replacing the viral enhancer. This element converts the ubiquitously active LTR into an erythroid enhancer-promoter, but the activity obtained with this vector is significantly lower than that of the wild type LTR (19, 21).

The present results show that simply adding a short sequence (HS1) to the GATA-1 HS2 and appropriately mutating a negatively acting element (Fig. 5) creates a much better cassette for hematopoietic expression that retains the hematopoietic specificity already shown for the HS2 cassette (19). The average activity conferred by the ΔHS1·HS2 element approaches that of the wild type LTR in both the SFCM (Fig. 3, *B* and *C*) and MΔN vectors (Fig. 6). The peak obtained with the ΔHS1·HS2 element is broader than that of SFCM, with a significant proportion of cells even exceeding the maximum expression level of the Moloney enhancer-based vector SFCM. The greater variance of the expression of the ΔHS1·HS2·SFCM vector *versus* the wild type SFCM vector is evidence that the former construct may not be fully insulated from position of integration effects. Nevertheless, the large majority of G418-resistant cells express the reporter gene, and the proportion of expressing cells remains fairly constant for tens of division cycles even in the absence of the selective agent. It will be of interest to test if these vectors can equally and efficiently drive globin gene expression in erythroid cells.

Acknowledgment—We thank Dr. G. Lamorte for help with FACS data analysis.

REFERENCES

1. Weiss, M. J., and Orkin, S. H. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 9623–9627
2. Yamamoto, M., Takahashi, S., Onodera, K., Muraosa, Y., and Engel, J. D. (1997) *Genes Cells* **2,** 107–115

3. Pevny, L., Lin, C. S., D'Agati, V., Simon, M. C., Orkin, S. H., and Costantini, F. (1995) *Development* **121,** 163–172
4. Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A., and Orkin, S. H. (1997) *EMBO J.* **16,** 3965–3973
5. Simon, M. C., Pevny, L., Wiles, M. V., Keller, G., Costantini, F., and Orkin, S. H. (1992) *Nat. Genet.* **1,** 92–98
6. Takahashi, S., Onodera, K., Motohashi, H., Suwabe, N., Hayashi, N., Yanai, N., Nabesima, Y., and Yamamoto, M. (1997) *J. Biol. Chem.* **272,** 12611–12615
7. McDevitt, M. A., Shivdasani, R. A., Fujiwara, Y., Yang, H., and Orkin, S. H. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 6781–6785
8. Tsai, S. F., Strauss, E., and Orkin, S. H. (1991) *Genes Dev.* **5,** 919–931
9. Nicolis, S., Bertini, C., Ronchi, A., Crotta, S., Lanfranco, L., Moroni, E., Giglioni, B., and Ottolenghi, S. (1991) *Nucleic Acids Res.* **19,** 5285–5291
10. Hannon, R., Evans, T., Felsenfeld, G., and Gould, H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 3004–3008
11. Cairns, L. A., Crotta, S., Minuzzo, M., Moroni, E., Granucci, F., Nicolis, S., Schiro, R., Pozzi, L., Giglioni, B., Ricciardi-Castagnoli, P., and Ottolenghi S. (1994) *EMBO J.* **13,** 4577–4586
12. Onodera, K., Takahashi, S., Nishimura, S., Ohta, J., Motohashi, H., Yomogida, K., Hayashi, N., Engel, J. D., and Yamamoto, M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 4487–4492
13. McDevitt, M. A., Fujiwara, Y., Shivdasani, R. A., and Orkin, S. H. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 7976–7981
14. Ronchi, A., Ciro, M., Cairns, L., Basilico, L., Corbella, P., Ricciardi-Castagnoli, P., Cross, M., Ghysdael, J., and Ottolenghi, S. (1997) *Genes Funct.* **1,** 245–258
15. Vyas, P., McDevitt, M. A., Cantor, A. B., Katz, S. G., Fujiara, Y., and Orkin, S. H. (1999) *Development* **12,** 2799–2811
16. Nishimura, S., Takahashi, S., Kuroha, T., Suwabe, N., Nagasawa, T., Trainor, C., and Yamamoto, M. (2000) *Mol. Cell. Biol.* **20,** 713–723
17. May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K. M., Luzzatto, L., and Sadelain, M. (2000) *Nature* **406,** 82–86
18. Pawliuk, R., Westerman, K. A., Fabry, M. E., Payen, E., Tighe, R., Bouhassira, E. E., Acharya, S. A., Ellis, J., London, I. M., Eaves, C. J., Humphries, R. K., Beuzard, Y., Nagel, R. L., and Leboulch, P. (2001) *Science* **294,** 2368–2371
19. Grande, A., Piovani, B., Aiuti, A., Ottolenghi, S., Mavilio, F., and Ferrari, G. (1999) *Blood* **93,** 3276–3285, and references therein
20. Moreau-Gaudry, F., Xia, P., Jiang, G., Perelman, N. P., Bauer, G., Ellis, J., Surinya, K. H., Mavilio, F., Shen, C. K., and Malik, P. (2001) *Blood* **98,** 2664–2672, and references therein
21. Lotti, F., Menguzzato, E., Rossi, C., Naldini, L., Ailles, L., Mavilio, F., and Ferrari, G. (2002) *J. Virol.* **76,** 3996–4007, and references therein
22. Sabatino, D. E., Seidel, N. E., Aviles-Mendoza, G. J., Cline, A. P., Anderson, S. M., Gallagher, P. G., and Bodine, D. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13294–13299, and references therein
23. Russell, D. (2001) *Blood* **98,** 2595–2596
24. Becker, P. B., Weih, F., and Schutz, G. (1993) *Methods Enzymol.* **218,** 568–587
25. Quivy, J. P., and Becker, P. B. (1993) *Nucleic Acids Res.* **21,** 2779–2781
26. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489
27. Ronchi, A., Nicolis, S., Santoro, C., and Ottolenghi, S. (1989) *Nucleic Acids Res.* **17,** 10231–10241
28. Singh, H., Sen, R., Baltimore, D., and Sharp, P. A. (1986) *Nature* **319,** 154–158
29. Verzeletti, S., Bonini, C., Marktel, S., Nobili, N., Ciceri, F., Traversari, C., and Bordignon, C. (1998) *Hum. Gene Ther.* **9,** 2243–2251
30. Cairns, L., Cirò, M., Minuzzo, M., Morlé, F., Starck, J., Ottolenghi, S., and Ronchi, A. (2003) *J. Cell. Physiol.* **195,** 38–49
31. Markowitz, D., Goff, S., and Bank, A. (1988) *J. Virol.* **62,** 1120–1124
32. Kinsella, T. M., and Nolan, G. P. (1996) *Hum. Gene Ther.* **7,** 1405–1413
33. Sanyal, A., and Schuening, F. G. (1999) *Hum. Gene Ther.* **10,** 2859–2868
34. Pikaart, M. J., Recillas-Targa, F., and Felsenfeld, G. (1998) *Genes Dev.* **12,** 2852–2862
35. Challita, M. P., and Kohn, D. B. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 2567–2571
36. Lorincz, M. C., Schubeler, D., Goeke, S. C., Walters, M., Groudine, M., and Martin, D. I. (2000) *Mol. Cell. Biol.* **20,** 842–850
37. Krall, W. J., Skelton, D. C., Yu, X. J., Riviere, I., Lehn, P., Mulligan, R. C., and Kohn, D. B. (1996) *Gene Ther.* **3,** 37–48
38. Phillips, K., Gentry, T., McCowage, G., Gilboa, E., and Smith, C. (1996) *Nat. Med.* **2,** 1154–1156