Single-Copy Transduction and Expression of Human γ-Globin in K562 Erythroleukemia Cells Using Recombinant Adeno-Associated Virus Vectors: The Effect of Mutations in NF-E2 and GATA-1 Binding Motifs Within the Hypersensitivity Site 2 Enhancer

By Jeffery L. Miller, Christopher E. Walsh, Paul A. Ney, Richard Jude Samulski, and Arthur W. Nienhuis

The use of recombinant adeno-associated virus (rAAV) vectors provides a new strategy to investigate the role of specific regulatory elements and trans-acting factors in globin gene expression. We linked hypersensitivity site 2 (HS2) from the locus control region (LCR) to a γ-globin gene (γ*) mutationally marked to allow its transcript to be distinguished from endogenous γ-globin mRNA. The vector also contains the phosphotransferase gene that confers resistance to neomycin (NeoR). HS2 region mutations within the NF-E2 motifs prevented NF-E2 binding while preserving AP-1 binding. Another set in the GATA-1 motif prevented binding of the factor. Several NeoR K562 clones containing a single unrearranged RAAV genome with the γ* gene linked to the native HS2 core fragment (WT), mutant NF-E2 HS2 (mut-NF2E), mutant GATA1 HS2 (mut-GATA1), or no HS ((-)HS) were identified. In uninduced K562 cells, mut-NF2E clones expressed γ* mRNA at the same level as the WT clones, compared with a lack of γ* signal in the (-)HS2 clones. However, hemin induction of mut-NF2E clones did not result in an increase in the γ* signal above the level seen in uninduced cells. Mut-GATA1 clones expressed the γ* mRNA at the same level as WT clones in both uninduced and induced cells. Thus, GATA-1 binding to this site does not appear to be required for the enhancing function of HS2 in this context. This single-copy rAAV transduction model is useful for evaluating the effects of specific mutations in regulatory elements on the transcription of linked genes. This is a US government work. There are no restrictions on its use.

Recombinant AAV (rAAV) have been shown to function as efficient gene transfer vectors. Only the ITR are required for encapsidation of the vector genome. The nonstructural and capsid proteins can be provided by a complementing helper genome from which the AAV ITR have been removed to prevent encapsidation. Cotransfection with plasmids containing the vector or helper genome and infection with wild-type adenovirus to provide its complementary functions results in release of rAAV and adenovirus. After heat-inactivation of the adenovirus, rAAV infect target cells efficiently and are integrated usually as 1 to 3 unrearranged tandem copies. Recently, a rAAV vector transduction system has been shown to provide high level expression of the γ-globin gene in K562 erythroleukemia cells. When linked to a wild-type HS2 enhancer core region, the γ* gene was transcriptionally active at levels of 40% to 110% of a single chromosomal γ-globin gene.

The goal of these experiments was to study the effects of mutations within the HS2 regulatory region on expression of the γ* gene using the rAAV vector transduction model to generate K562 erythroleukemia cell clones having a single unrearranged integrated vector genome. Previous analyses of HS function have generally relied on transfer methods that lead to multiple-copy integrants. With these methods, the effect of an individual regulatory element can only be inferred after correcting for gene copy number, and cooperative interactions between tandem gene copies are difficult to exclude. For example, in transgenic mice, single-copy versus multiple-copy transgenes appear to have separate requirements for LCR activity. We have shown that the rAAV vector system can be used as a novel strategy for the functional study of regulatory elements present as single-copy integrants.

MATERIALS AND METHODS

Generation of mutant HS2 DNA. Polymerase chain reaction (PCR)-directed site-specific mutagenesis was used to create the mutant HindIII-XbaI fragments shown in Fig. 1. pUC 007/HS2 served as the template. Inside primers included the substituted

From the Clinical Hematology Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD; and the Department of Biological Sciences, University of Pittsburgh, PA.

Submitted February 16, 1993; accepted May 24, 1993.

Address reprint requests to Arthur W. Nienhuis, MD, NIH, Blag 10, Rm 7C103, Bethesda, MD 20892.

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

This is a US government work. There are no restrictions on its use.

0006-4971/93/8206-0026$0.00/0

Blood, Vol 82, No 6 (September 15), 1993; pp 1900-1906
Hybonding.

Genomic DNA was digested with HincII and transferred to nitrocellulose.26 and mutant GATA-1 motif (mut-GATA1) erythroleukemia cells, and analysis of globin gene expression using bin gene sequences, stripped with 1% sodium dodecyl sulfate (SDS), and reprobed for the NeoR gene sequences.

sis on an agarose gel. The membranes were then probed for y-globin binding in gel-shift ouly.27 rAAV vectors containing wild-type HS2 (WT), no HS2 reverse transcriptase-PCR (RT-PCR) has been described previ-summer restriction sites of the human globin LCR. Protein binding regions including NF-E2 and GATA-1 motifs are indicated by boxes. Regions marked E and U denote protein binding from erythroid and ubiquitous cell extracts, respectively.3 (C) Mutated NF-E2 and GATA-1 protein binding motifs within the HS2 region. Two single-base mutations in the NF-E2 binding motif and paired triple-base mutations in the GATA-1 motif were made, as shown with arrows. The selected mutations prevent their respective protein binding in gel-shift assays.5,27

bases and outside primers included the HindIII and Xba I sites. VENT Polymerase (NE Biosci, Beverly, MA) was used in all PCR rounds (95°C for 1 minute, 55°C for 1 minutes, and 72°C for 1 minute, 30 cycles). Mutant fragments were then digested with HindIII and Xba I and subcloned into plasmid pSP72 (Promega, Madison, WI). Each mutant was then sequenced for structural confirmation using a USB Version 2.0 Kit (US Biochemicals, Cleveland, OH) before final construction of the pAAV/HS2/Ay*/NeoR vectors.

rAAV. Construction of recombinant virions, infection of K562 erythroleukemia cells, and analysis of globin gene expression using reverse transcriptase-PCR (RT-PCR) has been described previously.23 rAAV vectors containing wild-type HS2 (WT), no HS2 [(−)HS2], HS2 containing the mutant NF-E2 motif (mut-NFE2), and mutant GATA-1 motif (mut-GATA1) were assembled. All adenovirus and rAAV vector work was performed in a P-2 level laboratory.23

Nucleic acid analysis. Genome arrangement and copy number were determined by genomic DNA digestion and Southern blotting.27 Genomic DNA was digested with HindIII and transferred to Hybond N+ (Amersham, Arlington Heights, IL) after electrophoresis on an agarose gel. The membranes were then probed for γ-globin gene sequences, stripped with 1% sodium dodecyl sulfate (SDS), and reprobed for the NeoR gene sequences.

RNA isolation and analysis using RT-PCR was performed using PCR conditions, primers, and input RNA amounts identical to those previously described.10 The RT-PCR signals were used to calculate the level of γ* gene expression relative to endogenous γ-globin gene expression (Table 1). This method for quantitation of RNA expression from the transduced γ* gene has been shown to closely parallel values determined by RNase protection assays.10,23 Comparison of the mRNA-derived signals from the endogenous and transduced γ-globin genes was made using the densitometry function of a Molecular Dynamics phosphor-imager (Molecular Dynamics, Sunnyvale, CA).

RESULTS

Southern blot analysis. Ten single neomycin resistant colonies from each rAAV construct were isolated and analyzed for the presence of an intact HS2/γ* gene region and for single-copy integration (Fig 2). The γ-globin probe was used to identify DNA from clones that had an rAAV-transfected globin gene of proper insert size. The Neo probe, used to identify DNA representing rAAV-host genome junction fragments, confirmed integration and provided an accurate assignment of copy number. The criteria used to identify colonies of transduced K562 cells containing a single integrated rAAV genome were as follows. γ-Probed lanes hav-

**Table 1. Hemin-Induced γ* Globin Expression Relative to Endogenous γ-Globin Expression**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Pools (%)</th>
<th>Individual Clones (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT HS2</td>
<td>27 ± 6.3</td>
<td>Mean 66 ± 24.5†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. 75 ± 9.6†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. 78 ± 9.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. 29 ± 7.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. 81 ± 8.2</td>
</tr>
<tr>
<td>No HS2</td>
<td>≤5</td>
<td>Mean 7 ± 2.1 (P = .02)§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. 9 ± 3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. 8 ± 2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. ≤5</td>
</tr>
<tr>
<td>Mut-NFE2</td>
<td>8 ± 1.2 (P = .006) Mean 19 ± 6.0 (P = .04)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. 13 ± 4.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. 16 ± 4.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. 28 ± 3.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. 22 ± 4.1</td>
</tr>
<tr>
<td>Mut-GATA1</td>
<td>40 ± 30.8 (NS) Mean 97 ± 31.7 (NS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. 62 ± 8.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. 106 ± 13.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. 123 ± 6.3</td>
</tr>
</tbody>
</table>

Abbreviation: NS, not significant.

* Percentages reflect mean transduced γ* RT-PCR signal from the samples shown in Fig. 3 expressed as a percentage of the signal from a single endogenous γ-globin gene. The percentages were calculated as follows:

\[
\frac{\gamma^* \text{ RNA} \times \text{Endogenous Gene Copy No.}}{\gamma \text{ RNA} \times \gamma^* \text{ Gene Copy No.}} \times 100
\]

Signals at or near background level were assigned a lower limit of ≤5%.

† Mean and standard deviation of all determinations on the individual clones.

‡ The assay was performed in triplicate on the individual clones.

§ Statistical analyses were performed using a Wilcoxon's rank sum test.28 All P values reflect comparisons made with the WT HS2 group.
Fig 2. Southern blots of the integrated rAAV (HS2/γ*Neo⁰). (A) Schematic of the rAAV integrant. HincII digestion of the integrated rAAV genome results in two junction fragments and an HS2/γ*-containing insert. All blots were probed for the γ-globin sequences and then stripped and reprobed for the junction fragment containing Neo⁰ coding sequences. The γ-globin probe was derived by radiolabeling a 938-bp fragment flanked by Rsal and Xho restriction sites. The Neo probe was derived by radiolabeling a 600-bp fragment flanked by BglII and Nco restriction sites. (B through E) Southern blot analysis of single-copy stably integrated unrearranged rAAV containing clones. Diagrams show lanes probed with γ-globin (y) beside identical lane probed with the neomycin probe (N). All diagrams show a pool followed by single clones. γ-Probe shows an endogenous λγ band of 4 kb (○), an endogenous δγ band of 2.6 kb (●), and an HS2-containing band of 2.1 kb (→). Constructs without an HS2 region (C) showed a 1.7-kb λγ-containing band. The Neo probe showed junction fragments in all panels.

ing an λγ* band of correct size for the intact transferred λγ* gene and intensity of approximately one-third that of the endogenous bands were selected. Neo-probed signals from the selected lanes were checked for the presence of a single band representing a single junction fragment. Lanes containing bands other than those described were felt to represent nonclonal populations versus clonal populations having rearranged or multiple-copy integrants.

Pooled K562 populations were derived from 15 individual colonies isolated and combined after 15 days under neomycin selection. The pools were then hemin induced for 3 days before nucleic acid extraction and analysis. Southern blotting showed an average copy number of less than three in 5 pools from each HS2-containing construct (data not shown). Multiple junction fragments seen in the pools (Fig 2) confirm the presence of multiple clones within each population.

DNA from 40 single colonies was analyzed, with 10 for each vector construct. HincII-digested fragments of duplex rAAV DNA that have not integrated contain a Neo-probed
band of 1.9 kb; no bands of this length were seen (data not shown). Thus, the rAAV genome that conferred neomycin resistance had integrated into the K562 cell genome. As previously reported, a high frequency of clones (6 of 10) derived with the vector lacking the HS2 fragment had a rearranged AAV genome, possibly because this rAAV genome is only 80% the length of the wild-type AAV genome. Of the remaining 30 colonies picked for the other vector constructs, 15 had a single unrearranged rAAV genome as determined by the criteria established above, 4 had 2 or more copies of a unrearranged genome, 2 had a single rearranged genome, 4 had a combination of unrearranged and rearranged genomes, and 5 appeared to be mixture of more than one clone and thus were indeterminate.

**RNA expression of the transduced globin gene.** Three or four clones with a single unrearranged rAAV vector genome for each construct were grown and studied for the expression of the transduced γ-globin gene. RT-PCR signals from matched, uninduced and hemin-induced populations are shown in Fig 3. RT-PCR signals from the endogenous γ genes were hemin inducible in all pools and clones studied. An estimate of expression relative to a single endogenous γ-globin gene was made by multiplying the ratio of $^{\gamma*}\gamma/^\gamma$ by 6 to correct for the endogenous gene copy number. K562 cells are aneuploid for chromosome 1 (3 copies); hence, 6 endogenous γ-globin genes are present. Numerical expression data derived from the hemin-induced signals shown in Fig 3 are summarized in Table 1.

Analysis of RNA from mut-NF-E2 clones grown without hemin induction showed a level of $^{\gamma*}\gamma$ expression comparable to the WT HS2 containing uninduced clones (Fig 3C). This level of expression is greater than that seen in the uninduced clones without an HS2 region (Fig 3B). Therefore, HS2 without the NF-E2 binding motif is sufficient for uninduced expression of the transferred gene. In marked contrast to the signal generated by mRNA from the endogenous γ genes of these mutant NF-E2 clones, no appreciable increase in $^{\gamma*}\gamma$ expression was found with hemin induction. Clones containing the vector genome with the mut-GATA1 HS2 were also analyzed for $^{\gamma*}\gamma$ expression (Fig 3D). The pattern of expression mimicked that seen in the WT clones (Fig 3A). Therefore, mutation of a down-

![Fig 3. Detection and quantitation of $^{\gamma*}\gamma$ globin mRNA using RT-PCR. Molecular weight markers are shown on the left and controls on the far right. The RT-PCR product from the endogenous γ genes ($^\gamma$) and the 6 nt shorter RT-PCR product from the transduced marked $^{\gamma*}$ genes ($^{\gamma*}$) were generated using $^{32}$P-labeled dCTP in the PCR master mix. Five pools and three or four individual single-copy clones are shown for each construct. Individual panels (A through D) divide the results according to the HS2 region used in the rAAV constructs (see text for details). PCR products of RNA extracted from uninduced (U) and hemin-induced (I) K562 cells are shown in paired lanes. Matched (–) RT controls for all clones had no bands (data not shown).](image-url)
stream GATA-I binding motif did not interfere with the ability of HS2 to enhance γ-globin gene expression using this system. The variability in the level of expression among the pools was greater, as reflected by a larger standard deviation for mut-GATA1 HS2 than for WT HS2 (Table I), suggesting that the GATA-1 site may influence position-independent expression.

RT-PCR signals from the 20 pooled colonies described above are shown for comparison. The pattern of γ-globin gene expression seen in the pools was similar to that of the individual clones for each rAAV construct. The magnitude of hemin-induced γ-globin expression in the pools was approximately one-half that of individual clones (Table I). Correction for differences in average copy number and endogenous γ signal between pools are reflected in the mean expression values shown.

**DISCUSSION**

The rAAV vector system provides a novel strategy for achieving integration of a single unarranged gene with its transcriptional control elements. More than one-half of the K562 erythroleukemia cell clones derived byNeoR selection contained a single intact copy of the γ-globin gene linked to the HS2 from the LCR. Using the system, we verified that NF-E2 binding to the tandem NF-E2/AP-1 sites in HS2 is necessary for hemin inducibility, and we showed that the consensus GATA-1 binding motif within the HS2 core enhancer fragment is not required for high-level expression.

Our prior studies suggested that the rAAV vector system might be useful to obtain clones containing a low copy number of the vector genome. These new data provide a more precise quantitative estimate; 60% of 25 evaluable clones transduced with an rAAV having a genome 90% the length of the wild-type virus had a single unarranged copy and another 4 (16%) had 2 or more unarranged copies. The other 7 evaluable NeoR clones contained rearranged copies of the vector genome. Despite a modest difference in length between the vector lacking the HS2 fragment and the others, that vector gave a majority of NeoR clones (6 of 10) having a rearranged genome and 2 of the pooled populations derived with this vector had an average genome copy number of less than 0.5 (data not shown). We conclude that the rAAV system can be used for deriving clones having single-copy integrants but that each vector and perhaps each cell type will require characterization with respect to this capacity.

The LCR of the β-globin cluster has a number of activities that together result in high-level, regulated globin gene expression in matured erythroid cells.28,29 Establishment of each individual site most likely represents the displacement of one or two nucleosomes by sequence-specific binding of both erythroid-restricted and ubiquitous transcriptional factors. The individual sites are thought to interact in forming the active chromatin domain that encompasses the entire β-globin gene cluster in erythroid cells. These two functions, site formation and domain establishment, seem most likely to be the basis for the ability of the LCR or components thereof to confer position-independent, copy-number-dependent gene expression in transgenic animals. A third function, transcriptional enhancement, pertains specifically to the level of gene expression. In K562 erythroleukemia cells, the erythroid specific enhancer within HS2 has the greatest effect on the level of gene expression; whereas, in MEL cells and transgenic mice, HS3 and HS4 also confer high-level gene expression.30,31 Our experiments were designed specifically to test the enhancing function of HS2. Individual clones were selected based on expression of the linked NeoR gene; therefore, each integrated genome was likely to be in an active chromatin domain independent of HS2 function.

GATA-1 and NF-E2 motifs have been shown to be synergistic in creating a strong erythroid promoter when linked to a TATA box.32 The human γ-globin promoters contain GATA-1 binding motifs, but lack the NF-E2 motif. When a 46-bp fragment containing the tandem NF-E2/AP-1 sites from HS2 was linked to the γ-globin promoter, reporter activity in hemin-induced K562 cells was increased significantly.33,34 In contrast, when compared with the full HS2, the 46-bp fragment linked to the γ-globin promoter was not associated with high-level expression in stably transfected K562 cells.35 In MEL cells or transgenic mice, an additional 200 bp of the HS2, which includes the GATA-1 motif mutated in this study, were required for high-level expression of a linked globin gene.34,35

The nature of DNA sequences within the additional HS2 region that participate in high-level globin gene expression remains complex. Mutations of the GATA-1 motif in a synthetic 215-bp core HS2 fragment resulted in no reduction in the expression of a linked human β-globin when present as multiple-copy integrants in transgenic mice.24 The synthetic HS2 core did not function as a single copy in the transgenic mouse model. In pools of stably transfected MEL cells, a 38% reduction in β-globin expression was reported with the same GATA-1 mutated synthetic fragment. The effects of copy number on these pools was not reported. Also, independent of the GATA-1 mutation, the relative expression from those pools containing the wild-type synthetic HS2 was 75% less than the expression from those transfected with constructs having the HindIII-Xba I fragment of HS2.24 We were able to achieve high-level γ-globin expression from neomycin-resistant clones having a single copy of the rAAV genome that contained the full HindIII-Xba I fragment of HS2. Thus, the GATA-1 motif mutation was studied in clonal populations of K562 erythroleukemia cells independent of concatameric HS2 interactions. No mut-GATA1–related reduction in expression from the linked γ-globin was found using this model, although an influence on position-independent expression was suggested by the increased variability in expression level seen with the mut-GATA1 pools.

In vivo footprinting has been used to identify sites within HS2 occupied by proteins in intact K562 erythroleukemia cells.35,36 This methodology shows protein-DNA interactions in vivo and may identify functionally relevant binding motifs. Both the tandem NF-E2/AP-1 and GATA-1 HS2 sites are occupied in K562 erythroleukemia cells and the
pattern of protected nucleotides changes within each motif upon induction. NF-E2 appears to bind with induction, whereas GATA-1 may be released. Using the rAAV system to derive single-copy intergrants, we verified that the capacity for NF-E2 to bind to the tandem NF-E2/AP-1 sites is essential for the enhancing function of HS2. Binding of GATA-1, the only other erythroid-restricted transcriptional activator known to interact with this portion of HS2, did not appear to be required for inducible, high-level expression. These data are consistent with the results obtained by in vivo footprinting, and suggest the GATA-1 binding motif’s role in HS2 function in vivo may relate to effects on chromatin structure rather than transcriptional enhancement.

ACKNOWLEDGMENT

We thank Dr Jonathan Hibbs for providing the PCR-directed site-specific mutagenesis protocols, Dr Johnson M. Liu for plasmids PUC008 and pAAV/HS2/γ+/γ−/Neo and for reading the manuscript, and Jean Johnson for secretarial help in preparing the manuscript.

REFERENCES


11. Talbot D, Grosvenor F: The 5′HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. EMBO J 10:1391, 1991


Single-copy transduction and expression of human gamma-globin in K562 erythroleukemia cells using recombinant adeno-associated virus vectors: the effect of mutations in NF-E2 and GATA-1 binding motifs within the hypersensitivity site 2 enhancer [published erratum appears in Blood 1995 Feb 1;85(3):862]

JL Miller, CE Walsh, PA Ney, RJ Samulski and AW Nienhuis