Retargeting of Retroviral Integration Sites for the Predictable Expression of Transgenes and the Analysis of Cis-Acting Sequences[1]

Dirk Schübeler, Karin Maass, and Jürgen Bode*

GBF, National Research Center for Biotechnology, Mascheroder Weg 1, D-38124 Braunschweig, Germany

Received March 26, 1998; Revised Manuscript Received June 22, 1998

ABSTRACT: The transcriptional activity of transgenes in eukaryotic host cells critically depends on the sites of their integration where it is modulated by interactions between the promoter and surrounding chromatin structures. Retroviruses integrate their genome into chromosomal loci favoring expression from one long-terminal repeat (LTR). We have therefore developed a strategy in which retroviral vectors are provided with "tags", that is, targets for a site-specific recombinase (Flp). Presence of two 48 bp Flp recognition target (FRT) sequences permits the excision of a selection marker whereby the reading frame of a reporter gene (*lacZ*) is restored and β-galactosidase activity can be monitored to characterize the integration site regarding the level and stability of expression. The location of the remaining FRT tag permits the subsequent Flp-mediated insertion of a truncated selection marker which is then expressed from the LTR. This step represents an authentic site-specific recombination event which can be demonstrated by a number of criteria, among these its reversibility in the presence of Flp activity. Thereby the "expression trap" principle permits the efficient recovery of stable insertion events, and if a gene of interest is linked to the truncated marker, the established properties of a given genomic site can be utilized for transcription studies or for the generation of highly expressing clones, even for biotechnological purposes.

A recurrent problem in molecular biology and bioengineering is the low and variable activity of a transgene compared to certain endogenous counterparts. While work is in progress to define locus-control regions (LCRs) which actively open the structure of entire chromatin domains (*1*, *2*) or boundary sequences that insulate functional domains from each other (reviews: *3*−*5*), we have exploited the capability of retroviruses to select for integration sites with a high and stable transcriptional potential. In many instances the hallmark of these loci is a nearby DNAseI hypersensitive site (*6*) and/or sequences with an affinity for the nuclear matrix (*7*).

Recombinases of the λ integrase family such as Cre and Flp mediate site-specific recombination events at their specific target sequences (loxP or FRT,[1] respectively), and in principle, they can be applied to excise as well as to insert DNA in the genome of mammalian cells. Strategies in which retroviral vectors are used as vehicles to introduce a recognition target sequence at a favorable chromosomal site are therefore developed in order to guide transgenes to these predetermined loci (*8*−*11*). However, while recombinase-mediated excision represents an almost spontaneous monomolecular recombination event between two equally oriented target sequences, the reverse reaction (insertion) is significantly more complicated (*5*, *10*, *12*−*14*). Here we show that one of the requirements is a stringent selection system which efficiently prevents the expression of randomly integrated targeting constructs to enable the recovery of the desired event. To this end we have marked genomic sites with an FRT tag by retroviral infection and applied Flp recombinase to insert expression constructs carrying a corresponding tag. By this approach we are able to pose any transgene under the control of the retroviral 5′-LTR. We demonstrate that, regarding the level and stability of expression, the newly introduced gene follows the characteristics of the provirus. Since one of our cell lines (BHK) is widely used for the production of recombinant proteins (*15*), this system lends itself to the routine generation of highly expressing cell lines for a variety of biotechnological purposes. Apart from these obvious applications, the system also permits the analysis of cis-acting sequences at genomic reference integration sites. While standard gene-transfer techniques cause random and sometimes multiple copy integration events, which can only be evaluated if expression levels are averaged for many clones, the present system keeps the environment constant and allows the inspection of a single active copy. Any modulation of transcriptional characteristics can therefore unambiguously be attributed to an element's intrinsic features while the investigation of several distinct clones provides insights into the role of chromatin structure on its performance.

MATERIALS AND METHODS

Plasmids. The primary target, pMFNZ, was constructed by cutting pM5SePa (*7*) with *Sal*I/*Pst*I. The 3.5 kbp

† This work was supported by grants from Deutsche Forschungsgemeinschaft to J.B. (Bo 419/5-2; Bo 419/6-1) and by a career development grant from the Fonds der Chemischen Industrie to D.S.

* To whom correspondence should be addressed.

[1] Abbreviations: FACS, fluorescence-activated cell sorting; Flp, Flp recombinase; FRT, Flp recombinase target site; IRES, internal ribosomal entry site; LTR, long-terminal repeat.

S0006-2960(98)00705-3 CCC: $15.00
Published on Web 08/07/1998

fragment containing the MPSV (murine proliferative sarcoma virus)−LTRs and plasmid backbone was used to insert a 4.6 kbp PCR fragment derived from pNeoβgal (*16*) with a 5′ *Nsi*I site (compatible with *Pst*I) and a 3′ *Sal*I site containing the FRT*neo*FRT*lacZ* part of this vector.

Construction of the set of targeting vectors started with the generation of a 1.9 kbp PCR product from the plasmid TKNeo (provided by C. Karreman, GSF Munich) which contains a fusion gene of HSV-thymidine kinase and the neomycin resistance gene. The design of primers was chosen to produce a *tkneo*^{r} gene which is devoid of a start codon and delimited at its 5′ end by an FRT site in-frame with the gene. This PCR product was cut *Xba*I/*Hin*dIII and inserted into a 3.4 kbp *Xba*I/*Hin*dIII fragment from pMCLuPF2 (*17*). The resulting intermediate, pFTNP, was cut by *Sca*I, and the 2.6 kbp fragment was ligated to a 2.4 kbp *Sca*I fragment of pSBC-1 (*18*) and to a 4.2 kbp *Sca*I fragment of pMBCSeapLuc (*18*) to obtain pFTN and pFTNIL, respectively. Construct pFTN includes the ATG-free *tkneo*^{r} gene; pFTNIL contains a luciferase reporter gene in addition, linked by a poliovirus-derived internal ribosomal entry site (IRES) to *tkneo*^{r}. Details of the targeting constructs used for the S/MAR analyses (pFTNLuS) and the expression of Factor VII (pFTNI7) are described in ref *17*.

pFIG is another bicistronic vector carrying *flp* in the first cistron and the *gfp* in the second cistron. The construct was obtained using the pSBC system of Dirks et al. (*18*). The 1650 bp *flp* gene from pOG44 (*16*) was ligated into pSBC1 via the *Sma*I-*Sal*I sites to yield pSBC1FLP with *flp* under the control of the SV40 promoter. The *gfp* gene was excised from pGFP10 (*19*) and cloned into pSBC2 via the *Eco*RI and *Hin*dIII sites to yield pSBC2GFP. The *Xmn*I-*Not*I fragments from pSBC1FLP and pSBC2GFP were ligated to yield the functional construct pFIG (*20*).

Cell Culture and Transfection. NIH/3T3 murine embryo fibroblasts (ATCC CRL 1658), baby hamster kidney (BHK A (*21*)), a subclone of BHK-21 (ATCC CCL-10), and ψ-2 packaging cells (*22*) were cultured in DME medium containing 10% fetal calf serum, 20 mM glutamine, 60 μg/mL penicillin, and 100 μg/mL streptomycin and passaged at the time of confluence. Transfection routines were as described before (*23*) using calcium phosphate coprecipitation of circular plasmids.

Retroviral Gene Transfer. Infectious retroviral particles were generated after transfecting pMFNZ under stable expression conditions into ψ-2 packaging cells. Virus containing supernatant of these cells was used to infect the different target cells. After determining the multiplicity of infection (MOI) for each cell line, infection was performed under conditions that rendered one in a thousand cells resistant. Such a low titer reduces the risk of multiple infections in a given cell (*7*).

Flp-Mediated Reactions. Excision. Recombinase expression plasmid pFIG (*20*) was transfected, and 48 h later, Gfp expressing cells were isolated by FACS sorting.

Insertion. For gene targeting, the plasmid to be inserted was transfected together with an expression plasmid (pOG44) for Flp recombinase (*16*).

Reporter Assays. β-Galactosidase (FACS analysis). Viable cells were analyzed as described (*24*). Nonconfluent cells from a 9.1 cm^2 dish were washed with PBS, trypsinized, and suspended in PBS/10% FCS. After centrifugation (5 min, 200*g*) cells were dispersed in PBS/10% FCS at a concentration of 10^7/mL. One-hundred microliters of this suspension was transferred to a FACS tube, kept at 37 °C, and supplied with 100 μL of prewarmed (37 °C) solution of 2 mM fluoresceine-di-β(D)-galactoside; after 1 min the reaction was stopped by the addition of 2 mL of ice cold PBS/10% FCS. Cells were kept on ice until the test. FACS analyses were performed using an excitation wavelength of 488 nm.

β-Galactosidase (in Situ Staining). Cells were washed twice with PBS and covered with fixing solution (0.1% glutaraldehyde; 2% formaldehyde in PBS) for 2 min, washed with PBS, covered with staining solution [5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2mM $MgCl_2$, 100 μg/mL X-gal], and incubated overnight at 37 °C. After 20 h blue, β-galactosidase expressing clones were counted and referenced to the total number of clones.

Luciferase Test. The quantification of luciferase reporter activity was as described previously (*25*).

Northern Blots. For mRNA analysis (Figure 3B and 6D) cells from a 25 cm^2 culture dish were washed twice with PBS and harvested with TEN. The pellet was resuspended in 1 mL of Trizol (Gibco BRL) and then treated according to the manufacturer's protocol. Ten micrograms of total RNA was used per lane. Electrophoresis and hybridization were as described previously (*23*).

Southern Blots. Isolation of genomic DNA and Southern blot analysis followed standard procedures and are described in ref *12*.

RESULTS

Setting a Target. To exploit the favorable properties of retroviral integration sites, proviruses have to be established via the infection pathway. The retroviral vector used to this end should provide a suitable reporter gene which enables a characterization of these sites regarding their transcriptional properties and a tag for their potential reuse. Any identified suitable site can then serve as a target for the insertion of other constructs of choice.

The proviral vector applied in our strategy, pMFNZ, was derived from the murine proliferative sarcoma virus (MPSV), an ecotropic retrovirus which is widely used for transducing genes into various cell lines by infection (*26*). pMFNZ contains a *lacZ* transcription unit which at this stage is not expressed since its reading frame is interrupted by a *neo*^{r} selection cassette (*16*). The *neo*^{r} gene is driven by its associated promoter in the antisense direction relative to the proviral mRNA (Figure 1, MFNZ state), and this transcription unit is flanked by two equally oriented FRT sites allowing its excision by Flp recombinase later on.

After stable transfection into ψ-2 packaging cells, retroviral particles were recovered from the supernatant and used to infect different cell types. A low MOI was adjusted to obtain a single infection event per cell. G418-resistant clones were expanded and found by Southern blotting to harbor a single copy of the provirus (see below). The selection gene was then removed by a pulse of Flp recombinase posing the *lacZ* reporter under the control of the 5′-LTR. To this end cells were transfected with a construct (pFIG) mediating the simultaneous expression of Flp recombinase and green fluorescent protein (Gfp), the genes of which have been

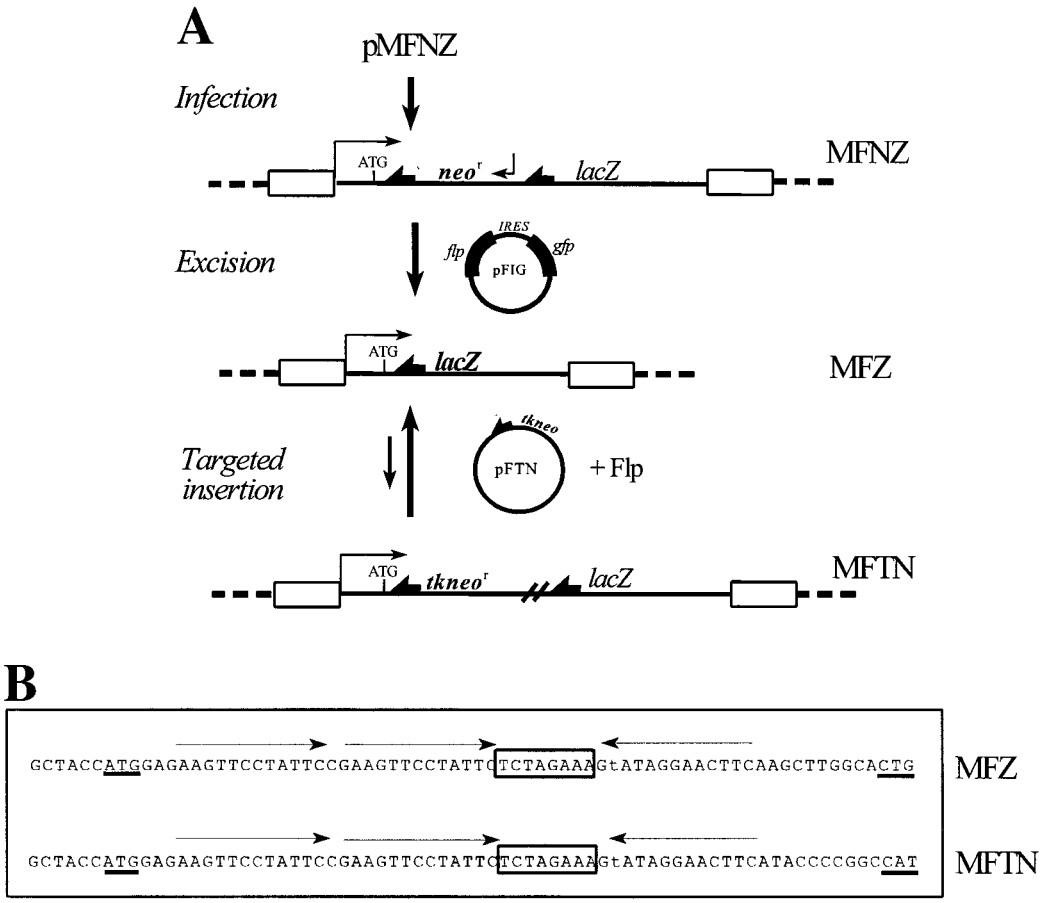

FIGURE 1: Establishment and reuse of genomic reference integration sites. (A) Infection. A proviral copy of MFNZ is introduced by first transfecting the vector in ψ-2 packaging cells, infecting cell lines with the resulting retroviral particles, and selecting for G418 resistance. Excision. Due to flanking FRT sites (half arrows) the *neo*^{r} gene can be excised by a pulse of Flp recombinase activity (generated by transient expression of pIFG; transfected cells sorted for Gfp expression). Thereby, the *lacZ* gene is posed under the control of the 5′-LTR (open box) promoter function and the ATG translation initiation codon. *lacZ* expressing cells are visualized by staining for β-galactosidase activity and are controlled for their expression level and stability at intervals over an extended time span. Targeted insertion. pFTN, a circular exchange vector carrying the *tkneo*^{r} fusion gene encoding thymidine kinase (HSV) and neomycin resistance (Tn5) activities, is inserted by site-specific recombination at the FRT site (half arrow). *tkneo*^{r} is a promoterless gene also lacking a translational start codon (ATG). Therefore only the targeted insertion should mediate resistance to G418. (B) Sequence and reading frame of MFZ [provirus after excision (*16*)] and the target construct after specific insertion. The start codon and the first codon of the lacZ reading frame (MFZ) and of *tkneo*^{r} (MFTN) are underlined. Arrows flanking the spacer (boxed sequence) indicate recombinase binding sites.

linked by an internal ribosomal entry site (IRES). Gfp positive cells were isolated by fluorescence-activated cell sorting (FACS) to enrich for those which have been successfully transfected and can therefore be expected to coexpress Flp. Subsequently, the sorted cells were stained for β-galactosidase activity since the excision event reconstitutes the *lacZ* reading frame (MFZ state in Figure 1) which can be visualized by an intracellular staining reaction. Using X-gal as the substrate, more than 90% of sorted cells displayed the "blue" phenotype compared to only 5% in the nonsorted control. These figures demonstrate that the sorting of transiently expressing cells strongly (20−50×) enriches for clones which have undergone the excision and that, in the presence of Flp recombinase, the reaction proceeds more or less spontaneously (*20*).

Next the precise expression characteristics of clones with a single provirus were established for a number of representative cell lines, among these BHK-A, a subclone of BHK 21 (*21*). The variation of β-galactosidase expressed per cell showed rather minor variations between different viral integration sites, that is, an up to 4-fold difference (Figure 2A). These data resemble a previous study on retroviral vectors with a different combination of reporter (SEAP-) and selector (PAC-) genes (*7*). All expression levels were rather constant over an extended period of time (six weeks, corresponding to 42 population doublings or 12 passages) as determined by β-galactosidase activity (Figure 2A) and single-cell expression analyses in a FACS assay (ref *24*; data not shown). Where tested, staining revealed no accumulation of nonexpressing clones even for time periods as long as three months.

Targeting the Provirus. The initial excision led to a provirus with a single remaining FRT site at a characterized genomic locus (MFZ state in Figure 1). In principle, such a provirus can serve as a target to accommodate different transgenes in order to make repeated use of its established properties. A complication for any insertion reaction of this type arises from the fact that it leads to a double-FRT structure (MFTN in Figure 1) which is easily lost by re-excision if the action of the recombinase does not terminate at the appropriate time (*10, 13*). In addition, if the targeting vector contains a functional selection gene, random integra-

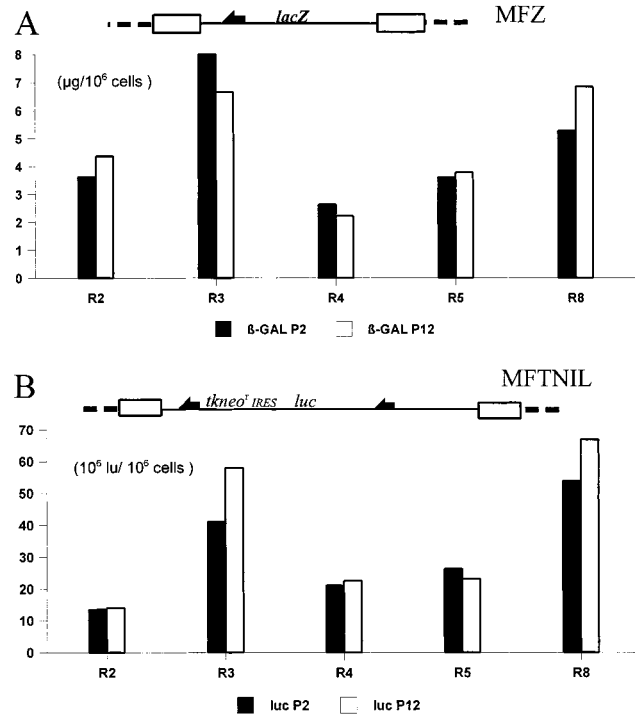

FIGURE 2: Expression characteristics and stability of clones derived according to Figure 1. (A) Five clones of the BHK cell line (R2−R8) were analyzed after excision of the *neo*^{r} gene for a single-copy infection event (cf. Figure 6), and their *lacZ* expression characteristics were established after two and twelve passages (P2, P12), respectively, that is, over a period of six weeks. (B) The same clones R2−R8 were used to insert a bicistronic *tkneo*^{r}*-luc* construct (pFTNIL) according to the principle delineated in Figure 1. Expression of the luciferase gene was again measured after two and twelve passages (P2, P12) and is seen to parallel *lacZ* expression.

tion of this construct leads to a high background of resistant but nontargeted clones (*9, 10*). The latter complication is particularly serious for the cell lines that are widely used because they are easy to transfect, meaning that they have a high capability of integrating DNA by illegitimate recombination.

Both limitations (excisional loss and random integration) will decrease the targeting efficiency. To reduce the background from randomly integrated but expressing copies, we decided to use a truncated selection marker as part of our targeting vector (pFTN). BHK cells are easily selected for G418 resistance, and therefore, the same selectable gene was used as for the infection step. This "selection marker recycling" concept (*27, 28*) which is enabled by site-specific recombinases, finds increasing use if successive manipulations are to be performed on a given cell clone.

For the insertion, the neomycin resistance function is encoded by a *tkneo*^{r} fusion gene which is devoid of a promoter and a translational start codon but which will be complemented by a Flp-mediated targeting into the proviral FRT site. At this position it will be expressed due to the pre-existing 5′-LTR and an in-frame ATG (MFTN state in Figure 1). Although the *tkneo*^{r} fusion gene in the targeting vector pFTN (Figure 1) and in its derivatives pFTNIL (Figure 2B) and pFTNI7 (Figure 3) contains an N-terminal extension, that is, the FRT sequence, its reading frame is maintained after insertion and the *tk*- as well as *neo*^{r} activities are preserved. At the same time the *lacZ* gene is interrupted

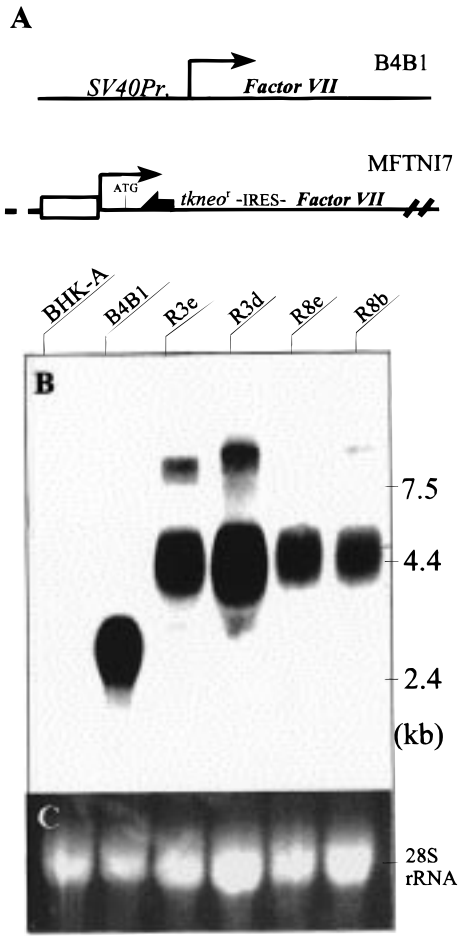

FIGURE 3: Expression of the human clotting factor VII gene at a proviral integration site. The principle used in Figure 2B for luciferase expression was applied for inserting the FVII gene at proviral integration sites R3 and R8. (A) Structure of the constructs used to express factor VII in the reference clone (B4B1) and in MFTNI7 after insertion into the proviral integration sites of R3 and R8. (B) mRNA levels, determined on a Northern blot with an *Eco*RI/*Mse*I probe from factorVII cDNA (4.2 kb band). Comparison with B4B1 (band at 2.4 kb) which had been obtained by selection and in which the gene is driven by an SV40 promoter. For the R3 and R8 subclones the band above the 7.5 kb marker (4%−7% of total intensity) results from polyadenylation at the 3′-LTR downstream from the SV40 polyadenylation signal next to the FVII gene. All experiments were in BHK cells (10 μg of total RNA/lane). (C) 28S rRNA signals after staining with ethidium bromide

enabling the facile detection of targeted clones by the blue-white screening principle introduced by O'Gorman et al. (*16*).

Applications of the Trap System. After characterizing the expression characteristics for a number of proviruses, we asked whether a vector sequence inserted in that position and posed under the control of the existing viral LTR is expressed at levels comparable to the initial *lacZ* reporter at the MFZ state (Figure 1). To this end we modified five of the above clones by a construct in which a luciferase gene is linked to the truncated selection marker by an IRES element and thereby coexpressed from the same mRNA (Figure 2B, MFTNIL state). Correctly targeted subclones were preselected by blue-white screening, then confirmed by Southern blot analysis (see below). For these clones the relative amount of luciferase activity is seen to follow the expression level of β-galactosidase in the initial proviruses (Figure 2B), suggesting that the influence of the integration site upon the viral promoter remains largely unaffected by

the nature of the coding region. Repeated measurements proved that the luciferase gene is expressed not only at a similar level but also with the same long-term stability as the *lacZ* gene prior to the insertion: even in the absence of selective pressure the expression was found to be stable for at least 12 passages as above (compare the pairs of bars in Figure 2A and 2B). This confirms our concept, the reuse of a characterized genomic site for the predictable expression of any gene.

The most highly expressing clones produced up to 8 μg of β-galactosidase/10^6 cells, that is, about 1% of total cellular protein (*29*). Luciferase was expressed at up to 6×10^7 light units (1.5 μg)/10^6 cells which may be due to its more complicated processing route (*30*), its lower stability (*31*), or its translation from the second cistron. Both β-galactosidase and luciferase were synthesized in amounts which are otherwise rather typical of transient transfection procedures (*32*). Considering that the BHK cell line is widely used for the production of recombinant proteins (*15*), our targeting protocol should be useful not only to obtain cell lines for experimental purposes but also for the rational production of therapeutically important proteins.

Factor VII as a Pharmaceutically Relevant "Gene of Interest". The best producers in this series, clones R3 and R8, were chosen for a direct comparison of their transcriptional potential with that of an established factor VII expressing BHK cell line. The underlying clone has been isolated at Novo Nordisk (Copenhagen) by a routine screening procedure of transfected clones and was used with kind permission. It was chosen because it harbors a single copy of the FVII c-DNA gene (C. Bornæs, private communication), in contrast to many other examples where high producers rely on multiple copy integration events which may cause long-term problems due to epigenetic inactivation events or genomic instability *(33, 34)*.

For this comparison we inserted a construct similar to pFTNIL but containing the FVII coding sequence (pFTNI7) to yield MFTNI7 (Figure 3A). Subclones of R3 and R8 with an authentic targeting event were identified by blue-white screening as well as by Southern blotting (see below), and four of these were assayed for transcriptional levels on a Northern blot (Figure 3). For the transfected FVII producer clone, the transcript is seen to migrate as a band slightly above 2.4 kb. Infected clones harboring pFTNI7 reveal a major transcript of about 4.2 kb as expected for a bicistronic mRNA which contains the selection marker in the first cistron and FVII cDNA in the second cistron. It is obvious that the amount of mRNA is comparable to the reference although a strict quantification of signal intensities has not been attempted due to the difference in mRNA length and to the fact that 28S rRNA migrates in the same range of the gel where it may cause interference with signal strength. Overall, these results support the potential of our targeting approach for the rational establishment of clones with a high transcriptional potential. They do not necessarily imply comparable levels of biological activity, as clotting factors undergo a highly complex postsynthetic processing pathway (*35*) which in a given cell may contain several bottlenecks beyond mRNA synthesis (C. Bornæs, personal communication).

Analysis of Cis-Acting Sequences at a Predefined Genomic Location. Conventionally, the study of transcriptionally active elements relies either on transient expression systems, which do not reconstitute the endogenous chromatin structure, or on large pools of stably transfected cells in which variations due to different integration sites and the number of active copies become averaged. Both approaches have their drawbacks, especially in the case of sequences which owe their activity to interactions with the chromosomal surroundings. This class of elements is exemplified by the scaffold- or matrix-attached regions (S/MARs) which require chromosomal integration for their function (*4*).

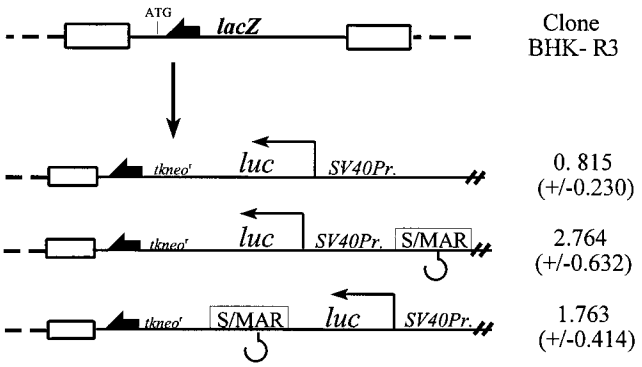

FIGURE 4: Transcriptional effect of a cis-acting element at a genomic reference integration site. Luciferase expression constructs with and without a S/MAR sequence were inserted into the genomic site of R3. Four independent subclones were characterized for luciferase activity to derive standard deviations.

On the basis of our previous work on S/MARs, we cloned an established 800 bp S/MAR fragment from the human interferon-β locus (specimen IV in ref *36*) into two different positions relative to a luciferase reporter. The reporter is located downstream from the selection trap and is controlled by its own SV40 promoter (Figure 4). There was no indication for a promoter occlusion due to interactions between the LTRs and the SV40 promoter, as transcription levels remained unchanged in the presence or absence of a continued G418 selection, at least for the time needed to conclude this experiment (4 weeks).

In previous experiments the 800 bp fragment and its parent S/MAR (I in ref *36*) augmented the expression level of a given promoter after stable integration but not in transient transfection assays which distinguished them from an enhancer (*23, 37*). After the reporter constructs were targeted into clone R3, the luciferase activity was determined for four derivative clones each. If positioned next to the promoter, the fragment augmented its activity about 3-fold, whereas at a position downstream from the gene the expression was approximately doubled. Since the subclones derived from the same targeting construct showed a relatively moderate variation ($\pm$25%), indicated by the standard deviations in Figure 4, this example underlines the use of our approach for the study of subtle transcriptional effects which cannot be addressed by standard transfection techniques. Moreover, this result is remarkable in that it reveals an additional transcriptional augmentation (*3, 4*) at an integration site which is already highly expressing (*7, 23*).

Performance of the Trap System. Figure 5 compiles data on the efficiency of the system in various cell lines and for three different constructs. It is seen that for pFTN and pFTNIL between 20% and 90% of all G418-resistant cells lost β-galactosidase activity because of the correct insertion.

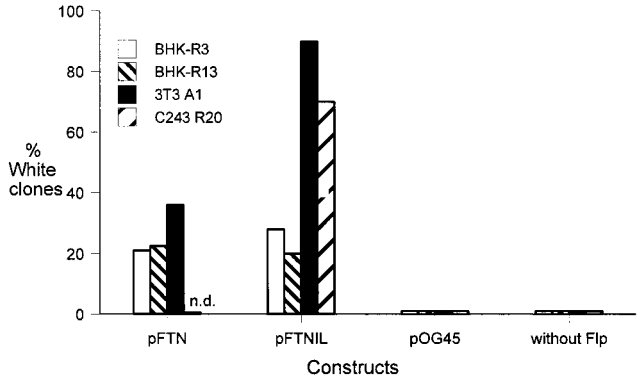

FIGURE 5: Efficacy of the expression trap in various cell clones. According to the Figure 1 principle, circular exchange vectors were transfected together with a recombinase expression plasmid into BHK-A, NIH/3T3, and C243 clones with established proviral targets. The percentage of white clones, that is, clones in which the *lacZ* gene had been separated from the control elements, indicates the proportion of correct insertion events. Efficiencies are compared for a basis construct, pFTN, containing only the *tkneo*^{r} cassette and for pFTNIL, a bicistronic construct containing the luciferase gene in addition (cf. Figures 1A and 2B). The targeting of pOG45 (*16*), a construct in which the *neo*^{r} gene is under the control of its own (CMV) promoter, yields no increased percentage of white clones relative to the control without Flp, demonstrating the efficiency of the selection trap. Bars labeled "without Flp" indicate the share of white clones in case the targeting constructs but no recombinase expression plasmid was used. This low percentage is ascribed to incomplete staining.

This result strictly depends on Flp expression since in its absence the share of white clones barely exceeds 1% (right-hand bars). The performance of the trap is demonstrated in an experiment with pOG45 (*16*), a complete vector which harbors a *neo*^{r} expression unit including an SV40 promoter. For this vector both random integration and Flp-guided targeting events will mediate G418 resistance. Since the percentage of white clones is around 1% as for the control in the absence of recombinase, the requirement of an expression trap for the cell lines used here becomes obvious.

The authentic nature of all recombination reactions was verified by Southern and Northern blotting procedures as exemplified by the insertion of targeting plasmid pFTNIL into one of the prominent BHK clones (R3; see Figure 6). The product of targeting was analyzed after *Sac*I (Sa) digestion for which site-specific insertion leads to a characteristic increase in fragment size (compare lanes 2 and 3 in Figure 6B). As predicted, this change is reversible since a new pulse of Flp activity excises the sequence intervening the FRT sites (Figure 6B, lane 4). Loss of the *tkneo*^{r} cassette leaves clones in a ganciclovir-insensitive state which was used to enrich this event by negative selection: the *tk* activity encoded by the *tkneo*^{r} fusion gene converts the drug into a cytotoxic nucleotide analogue which causes the elimination of all clones which have not undergone the excision.

Clones with a correct targeting event which confers G418 resistance were controlled for any additional, nonspecific integration of the targeting vector, and this was done by *Spe*I (Sp) digestion and probing with a *tk* fragment (II). Figure 6C shows an example of this control in which six out of eight clones yield only the signal for the desired Flp-mediated insertion whereas two of them harbor randomly integrated copies of the plasmid in addition to the authentic event (see extra bands in lanes 5 and 7). These clones were discarded

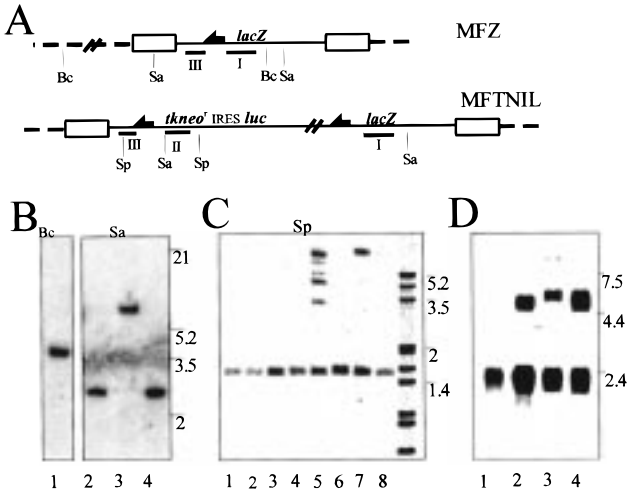

FIGURE 6: Control of authentic targeting and expression exemplified by clone R3. (A) Restriction sites and probes used to characterize the constitution of the provirus which is used as a target according to Figure 1: Bc, *Bcl*I; Sa, *Sac*I; Sp, *Spe*I. (B) Southern blots visualized by a *lacZ* probe (I). The presence of a single copy provirus is demonstrated by the generation of a unique bordering fragment, generated by *Bcl*I digestion of the parental clone genomic DNA (lane 1). An authentic insertion into this clone is demonstrated by the appropriate increase of the 2.5 kb *Sac*I fragment (lane 2) to 5.6 kb (lane 3); this event can be reversed by another pulse of Flp recombinase re-excising the targeting plasmid (lane 4). (C) Southern blot using a *tkneo* probe (II) to visualize additional random integrates. *Spe*I generates a defined 1.5 kb internal fragment in case of a Flp-mediated insertion of the *tkneo*^{r} construct (lanes 1−8). While all subclones yield the authentic 1.5 kb band, additional random integrations produce unique *Spe*I bordering fragments in case of lanes 5 and 7. (D) Northern blot visualized by the ψ-region probe III and a pyruvate kinase probe (internal standard). Lane 1: nonmanipulated BHK cells. Lane 2: infected cell clone (MFZ state, *lacZ* transcript). Lane 3: MFTNIL state (*tkneo*^{r}-*luc* transcript) following specific insertion. The minor band above 7.5 kb arises from termination within the 3′-LTR rather than at the luc-associated polyadenylation signal. Lane 4: same clone after Flp-mediated excision, reconstituting MFZ. Lanes 2−4 correspond to the events analyzed in part A.

to avoid any complication arising from other than the retargeted viral integration site. To establish the integrity of the proviral transcripts, Northern blots were prepared demonstrating the expected extension of the viral messenger RNA following a targeted insertion (see lane 3 in Figure 6D).

DISCUSSION

Specific targeting of well-characterized genomic sites can solve major problems associated with current gene-transfer techniques. Traditionally, these techniques tend to yield variable expression levels due to chromosomal position effects, and they may be severely affected by unpredictable cross interactions between the multiple copies which are usually integrated (*33*, *34*). If the vector is transferred as a circular plasmid, the site of its opening cannot be controlled. Moreover, for any transfected vector, whether it is initially circular or linear, the ends may become truncated during the integration process.

These complications can be overcome by using retroviral vectors in combination with a system for site-specific recombination. Retroviruses have the capacity to select sites on the basis of their transcriptional potential (*7*), and their

integration maintains the long-terminal repeats (LTRs). Except from the duplication of a few nucleotides of host DNA, there is no further rearrangement (ref *7* and references therein). If the retroviral vector carries a tag such as an FRT site, any construct of interest can be inserted at this site. This approach circumvents the repeated screening and characterization of clones which is otherwise required for each new task requiring the consistent expression of a transgene.

Our system has been developed for several purposes, among these the study of cis-acting elements at genomic reference integration sites. Such an approach is particularly relevant in cases where the action of an element is unpredictably affected by the number of integrated copies and by its interactions with the chromosomal surroundings (*5*). A prominent example are the S/MARs which, due to their affinity for the nuclear matrix, may direct a construct to a subclass of genomic sites which is not generally open for the S/MAR-free control. Some S/MAR constructs also tend to yield higher copy numbers due to their origin-of-replication activity and/or to their recombinogenic potential (*4, 5*). Both features are overcome by the tools we have combined here, that is, the retroviral integration machinery and a site-specific recombination system (see Figure 4).

The size of S/MAR-mediated transcriptional augmentation (factors 2 and 3, respectively) may appear small relative to those reported for standard transfection systems (refs *3*, *4*, and references therein), but it is comparable to the 5-fold augmentation in a situation where S/MARs are present in both the 5′- and the 3′-LTR (*7*). In this context several points have to be considered: (i) retroviruses integrate into transcriptionally active genomic regions with pre-existing S/MAR properties, and this may minimize the effect of an additional attachment region within the provirus (*7*); (ii) the transcriptional effect of such an additional S/MAR is highly dependent on its location relative to the promoter (*7*); and (iii) the sequences which are introduced between the LTRs by site-specific recombination may be affected (negatively or positively) by the regulatory elements residing in these regions.

Characterization and Reuse of Stable and Highly Expressing Sites. The strategy is based on a retroviral vector (pMFNZ) which contains a complete *neo*^{r} selection marker flanked by FRT sites (Figure 1). Since the marker interrupts the reading frame of a *lacZ* gene, its excision by Flp recombinase reconstitutes a provirus which will express the reporter at a site which, by definition, is (or has been) accessible to the enzyme. We have demonstrated that for all proviruses the expression varied no more than 4-fold and was stable over many passages of culture (Figure 2A). This stability contrasts observations on transfected constructs which become unstable with time due to silencing of the promoter, loss of the construct (Seibler, unpublished), or cosuppression phenomena (*33, 34*). We ascribe the properties of the proviruses to the mechanism of viral integration, since retrovirally encoded integrase is known to favor regions which are transcriptionally active (*38, 39*) and show a particular chromatin structure (*7*).

Although there are examples for the inactivation of endogenous retroviral promoters by methylation (*40*), inactivation of LTRs was not observed in this study. Here, the specific level of expression from the retroviral LTR was only determined by the nature of the integration site which, in the case of the BHK-derived clones, was rather uniform. Taking into account that for resistant cells the expression of the selection marker has to exceed a threshold, we cannot exclude that the rather low variation of activities is due to the design of the construct or even to the limited number of clones analyzed, although this seems unlikely if the similar results of a previous study on NIH/3T3 cells (*7*) are considered. In any event, our protocol allows the convenient recovery of clones which express the *lacZ* gene stably and at high rates. This is the prerequisite for a consistent expression of other genes at a given site or for comparing various constructs at suitable genomic locations.

Selection Trap Allows Efficient Targeting. For the cell types used here, Flp-mediated targeting events could only be isolated with a selection trap, that is, a resistance marker which is activated by insertion at the pre-existing FRT tag (Figure 5 and ref *10*). The use of a complete selection cassette would otherwise inevitably lead to a high background from nonspecific integration since these cell lines have a high transfection efficiency and thereby a considerable potential of illegitimate recombination. These observations agree with observations by Fukushige and Sauer (*41*) on the Cre/loxP system, although the efficiency of their trap system was not assessed by reference to a corresponding control.

This requirement of a trap is in apparent contradiction to the results of O'Gorman et al. (*16*) where the same construct used in Figure 5 (pOG45) yielded a high rate of specific insertion into a CV-1 clone. On the basis of this clone (E25B2) a test system has been developed which is successfully used in many laboratories including our own (*42*). In our opinion this discrepancy depends on the cell type studied, that is, the different capability to perform random integration as outlined above. As an example at the other extreme, embryonic stem (ES) cells are refractory to incorporating circular exchange constructs by other than site-specific recombination events, and therefore, they do not require precautions of the mentioned type (*12*).

Perspectives. The re-excision of a previously expressed transgene is the most stringent control for the cellular action of an encoded product (*43*). Our targeting vectors are derivatives of pFTN (Figure 1) which provides a fusion gene (*tkneo*^{r}) useful both for positive (*neo*^{r}) and for negative (*tk*) selection. In the presence of ganciclovir, HSV-Tk activity eliminates clones harboring the *tkneo*^{r} cassette and enables the enrichment of successful excision events. Excision leaves the host cell ready for remanipulation, interrupts the expression of the transgene, and permits a renewed application of the same selection marker.

The analysis of cis-acting sequences is desirable not only at their endogenous location but also at specific genomic reference sites with different characteristics. Retroviral infection leads to a set of (integration) sites which are well-suited for the expression of genes under the control of the LTR. These sites permit the researcher to investigate the performance of any (mutated) element under strictly comparable conditions revealing even subtle differences which only reflect the element's properties. We have performed a model experiment on the transcriptional effect of a S/MAR fragment. To this end the luciferase reporter gene was placed under the control of a separate promoter to uncouple its expression from the selection gene. Accordingly, luciferase expression does not have to follow the threshold imposed

onto the selection marker. In principle, this approach permits the assessment of any inducible and/or weak promoter element, which is rather difficult by standard transfection techniques.

The system described here has been optimized for the targeting of highly expressing sites in common producer cell lines. Inclusion of a site-specific recombination system enables the reuse of these sites and circumvents time-consuming screening procedures. It is anticipated that the establishment of production cell lines will be greatly facilitated by this concept since each clone is simply a variant of a predecessor with already known properties.

NOTE ADDED IN PROOF

In agreement with our studies (Figures 3 and 4 in ref *23*) Argarwal et al. (*44*) recently used the 800 bp S/MAR fragment to improve expression characteristics of a (therapeutical) retrovirus vector.

ACKNOWLEDGMENT

We appreciate the gift of a FVII expression construct and cell clone by Peter Müller (GBF; with the approval of Novo Nordisk, Copenhagen) and many stimulating discussions with our colleagues Michaela Iber and Jost Seibler at GBF as well as Ruth Simeson, Karen Hansen, Claus Bornæs, and Leif Kongerslev at Novo Nordisk. We thank Maria Höxter (GBF) for cell sorting and FACS analyses.

REFERENCES

1. Ellis, J., Tan-Un, K. C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S., and Grosveld, F. (1996) *EMBO J. 15*, 562−568.
2. Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M., and Kioussis, D. (1996) *Science 271*, 1123−1125.
3. Bode, J., Schlake, T., Ríos-Ramírez, M., Mielke, C., Stengert, M., Kay, V., and Klehr-Wirth, D. (1995) in *Structural and Functional Organization of the Nuclear Matrix − International Review of Cytology* (Jeon, K. W., and Berezney, R., Eds.) Vol. 162A, Chapter 8, pp 389−453, Academic Press, Orlando, FL.
4. Bode, J., Stengert-Iber, M., Schlake, T., Kay, V., and Dietz-Pfeilstetter, A. (1996) *Crit. Rev. Eukaryotic Gene Expression 6*, 115−138.
5. Bode, J., Bartsch, J., Boulikas, T., Iber, M., Mielke, C., Schübeler, D., Seibler, J., and Benham, C. (1998) *Gene Ther. Mol. Biol. 1*, 551−880.
6. Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987) *J. Virol. 61*, 336−343.
7. Mielke, C., Maass, K., Tümmler, M., and Bode, J. (1996) *Biochemistry 35*, 2239−2252.
8. Choulika, A., Guyot, V., and Nicolas, J. F. (1996) *J. Virol. 70*, 1792−1798.
9. Karreman, S., Hauser, H., and Karreman, C. (1996) *Nucleic Acids Res. 24*, 1616−1624.
10. Schübeler, D., Mielke, C., and Bode, J. (1997) *In Vitro Cell. Dev. Biol.: Anim. 33*, 825−830.
11. Vanin, E. F., Cerruti, L., Tran, N., Grosveld, G., Cunningham, J. M., and Jane, S. M. (1997) *J. Virol. 71*, 7820−7826.
12. Seibler, J., Schübeler, D., Fiering, S., Groudine, M., and Bode J. (1998) *Biochemistry 37*, 6229−6234.
13. Logie, C., and Stewart, A. F. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5940−5944.
14. Kilby, N. J., Snaith, M. R., and Murray, J. A. H. (1993) *Trends Genet. 9*, 413−421.
15. Spier, R. E., Griffith, J. B., and Meignier, B., Eds. (1991) *Production of Biologicals from Animal Cells in Culture*, Butterworth-Heinemann, Oxford, U.K.
16. O'Gorman, S., Fox, D. T., and Wahl, G. M. (1991) *Science 251*, 1351−1355.
17. Schübeler, D. Ph.D. Thesis, University of Braunschweig, 1998.
18. Dirks, W., Wirth, M., and Hauser, H. (1993) *Nucleic Acids Res. 128*, 247−249.
19. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992) *Gene 111*, 229−233.
20. Iber, M. Ph.D. Thesis, University of Braunschweig, 1997.
21. Karreman, C., Karreman, S., and Hauser, H. (1996) *Virology 220*, 46−50.
22. Mann, R., Mulligan, R. C., and Baltimore, D. (1983) *Cell 33*, 153−159.
23. Schübeler, D., Mielke, C., Maass, K., and Bode, J. (1996) *Biochemistry 35*, 11160−11169.
24. Fiering, S. N., Roederer, M., Nolan, G. P., Micklem, D. R., Parks, D. R., and Herzenberg, L. (1991) *Cytometry 12*, 291−301.
25. Seibler, J., and Bode, J. (1997) *Biochemistry 36*, 1740−1747.
26. Artelt, P., Morelle, C., Ausmeier, M., Fitzek, M., and Hauser, H. (1988) *Gene 68*, 213-219.
27. Cregg, J. M., and Madden, K. R. (1989) *Mol. Gen. Genet. 219*, 320−323.
28. Abuin, A., and Bradley, A. (1996) *Mol. Cell. Biol. 16*, 1851−1856.
29. Patterson, M. K. (1979) *Methods Enzymol. 58*, 141−152.
30. Gould, S. J., Keller, G.-A., and Subramani, S. (1987) *J. Cell Biol. 105*, 2923−2931.
31. Alam, J., and Cook, J. L. (1990) *Anal. Biochem. 188*, 245−254.
32. Paborski, L. R., Fendly, B. F., Fisher, K. L., Lawn, R. M., Marks, B. J., McCray, G., Tate, K. M., Vehar, G. A., and Gorman, C. M. (1990) *Protein Eng. 3*, 547−553.
33. Dorer, D. R. (1997) *Transgenic Res. 6*, 3−10.
34. Wolffe, A. P. (1998) *Nat. Genet. 18*, 5−6.
35. Nakagaki, T., Foster, D. C., Berkner, K. L., and Kisiel, W. (1991) *Biochemistry 30*, 10819−10824.
36. Mielke, C., Kohwi, Y., Kohwi-Shigematsu, T., and Bode, J. (1990) *Biochemistry 29*, 7475−7485.
37. Klehr, D., Maass, K., and Bode, J. (1991) *Biochemistry 30*, 1264−1270.
38. Hwang, L.-H. S., and Gilboa, E. (1984) *J. Virol. 50*, 417−424.
39. Scherdin, U., Rhodes, K., and Breindl, M. (1990) *J. Virol. 64*, 907−912.
40. Groudine, M., Eisenmann, R., and Weintraub, H. (1981) *Nature (London) 292*, 311−317.
41. Fukushige, S., and Sauer, B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7905−7909.
42. Schlake, T., and Bode, J. (1994) *Biochemistry 33*, 12746−12751.
43. Westerman, K. A., and Leboulch, P. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 8971−8976.
44. Agarwal, M., Austin, T. W., Morel, F., Chen, J. Y., Bohnlein, E., and Plavec, I. (1998) *J. Virol. 72*, 3720−3728.

BI9807052