

High-level transgene expression by homologous recombination-mediated gene transfer

Mélanie Grandjean¹, Pierre-Alain Girod¹, David Calabrese¹, Kaja Kostyrko¹, Marianne Wicht², Florence Yerly³, Christian Mazza³, Jacques S. Beckmann², Danielle Martinet² and Nicolas Mermod^{1,*}

¹Laboratory of Molecular Biotechnology, Center for Biotechnology UNIL-EPFL, University of Lausanne,

²Service of Medical Genetics, Center Hospitalier Universitaire Vaudois (CHUV), 1015 Lausanne and

³Department of Mathematics, University of Fribourg, 1700 Fribourg, Switzerland

Received August 24, 2010; Revised May 12, 2011; Accepted May 13, 2011

ABSTRACT

Gene transfer and expression in eukaryotes is often limited by a number of stably maintained gene copies and by epigenetic silencing effects. Silencing may be limited by the use of epigenetic regulatory sequences such as matrix attachment regions (MAR). Here, we show that successive transfections of MAR-containing vectors allow a synergistic increase of transgene expression. This finding is partly explained by an increased entry into the cell nuclei and genomic integration of the DNA, an effect that requires both the MAR element and iterative transfections. Fluorescence *in situ* hybridization analysis often showed single integration events, indicating that DNAs introduced in successive transfections could recombine. High expression was also linked to the cell division cycle, so that nuclear transport of the DNA occurs when homologous recombination is most active. Use of cells deficient in either non-homologous end-joining or homologous recombination suggested that efficient integration and expression may require homologous recombination-based genomic integration of MAR-containing plasmids and the lack of epigenetic silencing events associated with tandem gene copies. We conclude that MAR elements may promote homologous recombination, and that cells and vectors can be engineered to take advantage of this property to mediate highly efficient gene transfer and expression.

INTRODUCTION

A major impediment to efficient and stable transgene expression is the variability of expression noted in independently transformed mammalian cells and organisms, both in experimental biology and for therapeutic applications. The high degree of expression variability is thought to depend on the number of transgene copies that integrate within the host genome and on the site of transgene integration (1,2). Indeed, transgene expression may be influenced by the fortuitous presence of regulatory elements at the random integration locus in the host genome. In addition, transgene expression is thought to reflect the influence of particular chromatin structure coming from adjacent chromosomal domains (3–5). Finally, the co-integration of multiple transgene copies at the same genomic locus may lead to silencing, possibly because of the formation of small inhibitory RNAs from antisense transgene transcription (6).

To increase and stabilize transgene expression in mammalian cells, epigenetic regulators such as matrix attachment regions (MAR) are increasingly used to protect transgenes from silencing effects (7). MAR were first discovered two decades ago for their association with the nuclear matrix or scaffold (8,9), a poorly characterized structural network that may consist of various non-histone nuclear proteins such as lamins, topoisomerases and components of transcription machinery (10). Eukaryotic chromosomes are organized in independent loops of chromatin that may control DNA replication, transcriptional regulation and chromosomal packaging (11–15). MARs were proposed to be the specific DNA sequences that anchor the chromosomes to the matrix

*To whom correspondence should be addressed. Tel: +41 21 693 61 51; Fax: +41 21 693 76 10; Email: nicolas.mermod@unil.ch
Present addresses:

Pierre-Alain Girod, Selexis SA, 18 Ch. des Aulx, Geneva, Switzerland.

David Calabrese, Selexis SA, 18 Ch. des Aulx, Geneva, Switzerland.

and partition chromosomes into these 50–200 kb DNA loop structures (16–18).

MARs are polymorphic 300–3000 bp-long DNA elements composed essentially of non-coding AT-rich sequences, and they are estimated to be 50 000–100 000 in the mammalian genomes (10). Their activity is thought to relate to their structural properties rather than to their primary sequence. Although no consensus MAR sequence has been found, they often have AT-rich sequences (19) and they may adopt particular conformations and physicochemical properties, such as a natural curvature (20), a deep major groove and a narrow minor groove (21), a high DNA strand unwinding and unpairing susceptibility (12), and a high potential to double-helix denaturation (22,23).

Besides providing a topological structure to the chromatin, MARs also contribute to regulate key genomic functions (24), as they were involved in the control of activities such as DNA replication and gene transcription (25,26). For instance, several origins of replication have been mapped within MARs in various eukaryotic genomes (27). Moreover, MARs are able to recruit endogenous replication factors and may allow sustained episomal replication when placed within an active transcription unit (28,29). Similarly, the ability of MARs to influence gene expression has been associated to the binding of protein factors in addition to the intrinsic properties of their DNA sequence (8,30,31). MARs associate with specific ubiquitous and tissue-specific transcription factors such as special AT-rich binding protein1 [SATB-1; (32)], NMP4 (33) and CTCF (34), which may in turn recruit regulatory proteins such as histone acetyl transferases, topoisomerases and ATP-dependent chromatin remodeling complexes to mediate a more expression-permissive chromatin state (35,36), as well as components of the transcription machinery and splicing factors (37,38). Thus, in addition to defining chromatin loop domains and organizing chromosomal architecture, MARs may contribute to control chromatin structure and gene expression.

MAR elements were shown to increase transgene expression and to decrease the clonal variability in stable transfections of mammalian cell lines and in transgenic plants and animals (21,39–42). MARs were proposed to act as insulators that protect transgenes from the repressive effects of surrounding heterochromatic area of the chromosomes and/or to relocate transgenes in an active compartment of the nucleus (7). MAR were also incorporated into viral vectors to reduce their susceptibility to silencing (43). A transgene flanked with MAR elements may thus constitute an autonomous chromatin domain whose expression would remain independent of the adjacent chromosomal environment.

MAR elements may also constitute targets of DNA recombination or rearrangement events, as exemplified by many MAR-related deletions and translocations involved in leukaemia and breast cancer (44,45). Additionally, it was shown that retroviral integration often occurs close to or within MARs (46,47). MARs were shown to bind DNA topoisomerase II (48), an enzyme that catalyzes double-strand breaks, as may be required to initiate the

recombination pathways of DNA repair (49,50). However, whether MARs might mediate or regulate DNA recombination remains to be demonstrated.

In this study, we show that successive transfections of MAR-containing expression vectors mediate an unexpectedly high increase in transgene expression when the DNA reaches the nucleus at phases of the cell division cycle when the homologous recombination (HR) pathway is most active. This results in part from an increased transport of the DNA in the nucleus and more efficient genomic integration. This effect is abrogated in cells deficient in HR as opposed to non-homologous end-joining. This study thus allows to propose a new function of MAR elements, which may be to promote homologous DNA recombination. It also identifies some of the limitations to efficient gene transfer and expression in eukaryotic cells, and it provides new avenues for more efficient and more reliable gene expression, for instance to express therapeutic proteins or for gene and cell therapies.

MATERIALS AND METHODS

Plasmids and constructs

pGEGFPcontrol contains the SV40 early promoter, enhancer and vector backbone from pGL3 (Promega) driving the expression of the eGFP gene from pEGFP-N1 (Clontech). pPAG01SV40EGFP results from the insertion of the chicken lysozyme MAR fragment upstream of the SV40 early promoter of pGEGFPcontrol (51). The human MAR 1–68 was identified by the SMARScan program using DNA structural properties. It was cloned from human bacterial artificial chromosomes in pBluescript and then inserted into pGEGFPcontrol upstream the SV40 early promoter, resulting in the p1–68(NcoI filled)SV40EGFP plasmid (21).

pGL3-CMV-DsRed was created by inserting the DsRed gene, under the control of the CMV promoter (including the enhancer), from pCMV-DsRed (Clontech) in pGL3-basic (Promega). pGL3-CMV-DsRed-kan was then created by exchanging the ampicillin gene of pGL3-CMV-DsRed for kanamycin resistance gene from pCMV-DsRed (Clontech) by digestion of both plasmids with BspHI. Then, the chicken lysozyme or the human 1–68 MAR were inserted into the pGL3-CMV-DsRed-kan plasmid. They were inserted as KpnI/BglII fragment containing the chicken lysozyme fragment, or as KpnI/BamHI human 1–68MAR fragment, upstream of the CMV promoter in pGL3-CMV-DsRed-kan, resulting in pPAG01GL3-CMV-DsRed and p1–68(NcoI) filledGL3-CMV-DsRed, respectively.

Cell culture and transfection

The CHO DG44 cell line (52) was cultivated in DMEM: F12 (Gibco) supplemented with HT (Gibco) and 10% fetal bovine serum (FBS, Gibco). Parental CHO cells AA8, NHEJ deficient cells V3.3 (53) and HR deficient cells 51D1 (54) were kindly provided by Dr Fabrizio Palitti (University of Tuscia, Italy) and were cultivated in DMEM: F12 medium with 10% FBS and antibiotics.

Transfections were performed using Lipofect-AMINE 2000, according to the manufacturer's instructions (Invitrogen). GFP or DsRed fluorescence levels were analyzed using a fluorescence-activated cell sorter (FACS), 1, 2 or 3 days post-transfection (transient transfections). Stable pools of CHO-DG44 cells expressing GFP and/or DsRed were obtained by cotransfection of the resistance plasmid pSVpuro (Clontech). After two weeks of selection with 5 µg/ml puromycin for CHO-DG44 (8 µg/ml puromycin for AA8, V3.3 and 51D1), cells were analyzed by FACS.

Cell subjected to multiple transfections were treated as follows: after the first transfection, the cells were transfected a second time according to the protocol detailed above, except that the resistance plasmid carried a different resistance gene (pSV2 neo, Clontech). The two transfections were at a 21 h interval, to be in phase with the cell cycle progression, unless otherwise indicated in the text. Twenty-four hours after the second transfection, cells were passaged and selected with 250 µg/ml G418 and/or 2.5 µg/ml puromycin (250 µg/ml G418 and 4 µg/ml puromycin for AA8, V3.3 and 5A1D1). After 3 weeks of selection, GFP and/or DsRed expression was analyzed by cytofluorometry.

Fluorescence activated cell sorting

Transient expression of eGFP and DsRed proteins was quantified at 24, 48 or 72 h after transfection using a FACScalibur flow cytometer (Becton Dickinson), whereas expression of stable cell pools was determined after at least 2 weeks of antibiotic selection. Cells were washed with PBS, harvested in trypsin-EDTA, pooled, and resuspended in serum-free synthetic ProCHO5 medium (Cambrex corporation). Fluorescence analyses were acquired on the FACScalibur flow cytometer (Becton Dickinson) with the settings of 350 V on the GFP channel (FL-1) and 450 V on the DsRed channel (FL-3) for transient expression. Settings of 240 V for FL-1 and 380 V for FL-3 were used to analyze stable expression. 100 000 events were acquired for stable transfections and 10 000 for transient transfections. Data processing was performed using the WinMDI 2.8 software.

Cell cycle analysis

At the indicated times, the cell cycle status was analyzed by flow cytometry after staining the nuclear DNA with propidium iodide (PI). Cells were first washed with a (PBS), trypsinized and harvested in 1 ml of growth media by centrifugation for 5 min at 1500 rpm (250 g) in a microcentrifuge. After an additional PBS wash, cells were resuspended in 1 ml of PBS before fixing with ethanol by the addition of 500 µl of cold 70% ethanol dropwise to the cell suspension under agitation in a vortex. Samples were incubated for 30 min at -20°C and cells were centrifuged as before. The resulting cell pellet was resuspended in 500 µl of cold PBS, supplemented with 50 µg/ml of RNaseA and the DNA was stained with 40 µg/ml of PI for 30 min in the dark. Cells were then washed with PBS, centrifuged and resuspended in 500 µl of ProCHO5 medium (Cambrex corporation) before

analysis in a FACScalibur flow cytometer (FL-3 channel; Becton Dickinson). Ten thousand events were acquired for each sample.

Fluorescent in situ hybridization

Fluorescent *in situ* hybridization (FISH) were performed as previously described in Derouazi *et al.* (55) and Girod *et al.* (21). Briefly, metaphase chromosomal spreads were obtained from cells transfected with or without the 1-68 human MAR and treated with colchicine. FISH was performed using hybridization probes prepared by the direct nick translation of pSV40GEGFP plasmid without the MAR.

Isolation of nuclei and DNA

Nuclei were isolated 1, 2 or 3 days after transient transfection(s), from proliferating and confluent CHO DG44 cells grown in 6-well plates. Cells (1×10^6) were washed twice with cold PBS, resuspended in 2 volumes of cold buffer A [10 mM HEPES (pH 7.5), 10 mM KCl, 1.5 mM Mg(OAc)₂, 2 mM dithiothreitol] and allowed to swell on ice for 10 min. Cells were disrupted using a Dounce Homogenizer. The homogenate was centrifuged for 2 min at 2000 rpm (370 g) at 4°C. The pellet of disrupted cells was then resuspended in 150 µl of PBS and deposited on a cushion of buffer B [30% sucrose, 50 mM Tris-HCl (pH 8.3), 5 mM MgCl₂, 0.1 mM EDTA] and centrifuged for 9 min at 3500 rpm (1200 g). The pellets of nuclei were resuspended in 200 µl of buffer C [40% glycerol, 50 mM Tris-HCl (pH 8.3), 5 mM MgCl₂, 0.1 mM EDTA] and stored frozen at -80°C until use (56).

Total cell DNA was isolated from CHO DG44 stable cell pools or from isolated cell nuclei using the DNeasy Tissue Kit from Qiagen. For stable cell pools, 1×10^6 confluent CHO DG44 cells growing in 6-well plates were collected. DNA extraction was performed according to the manufacturer's instruction for the isolation of total DNA from cultured Animal cells. DNA isolation was performed on frozen pellets of isolated cell nuclei, which were first thawed and centrifuged at 1700 rpm (300 g) for 5 min to remove buffer C before beginning DNA extraction following the same protocol as for the isolation of DNA from stable cell lines.

Transgene copy number determination and quantitative PCR

To determine the copy number of transgenes integrated in the genome, ~6 ng of genomic DNA were analyzed by quantitative PCR using the SYBR Green-*Taq* polymerase kit from Eurogentec Inc and ABI Prism 7700 PCR machine. The following primers were used to quantify the GFP gene: GFP-For: ACATTATGCCGACAAA GCC and GFP-Rev: TTGTTTGGTAATGATCAGCAA GTTG, while primers GAPDH-For: CGACCCCTTCAT-TGACCTC and GAPDH-Rev: CTCCACGACATACTC AGCACC were used to amplify the GAPDH gene. The ratios of the GFP target gene copy number were calculated relative to that of the GAPDH reference gene as described previously (57). To determine import of the transgene into nuclei following transfection, quantitative

PCR was performed on DNA extracted from purified nuclei using the same GFP and GAPDH primer pairs as described above.

The number of GAPDH gene and pseudogene copies used as reference was estimated for the mouse genome, as the CHO genome sequence is not available as yet. Alignment to the mouse genome of the DNA sequence of the 190 bp amplicon generated by the GAPDH primers was performed using NCBI BLAST software. A number of 88 hits was found per haploid genome. As the CHO DG44 are near-diploid cells (55), we estimated that 176 copies of the GAPDH genes and pseudogenes occur in the genome of CHO DG44 cells. This number was used as a normalization standard for the quantification of the GFP transgene copy number.

Confocal microscopy

pGEGFPcontrol and p1-68(NcoI filled)SV40EGFP plasmids were labeled either with rhodamine by the *Label IT Tracker TH-Rhodamine Kit* or with Cy5 by the *Label IT Tracker Cy 5 Kit* (Mirus, Mirusbio) according to the manufacturer's protocol, and purified by ethanol precipitation. For transfection, DNA transfection was carried out with the Lipofectamine 2000 reagent (Invitrogen) according to the supplier's instructions. At 3, 6 and 21 h after transfection, the medium was removed and the cells were fixed with 4% paraformaldehyde at room temperature for 15 min. When indicated, cells were treated for 30 min with LysoTracker Red DND-99 (Molecular Probes, Invitrogen) at a final concentration of 75 nM before fixation, to stain the acidic organelles (e.g. endosomes and lysosomes) according to the manufacturer's instructions. The fixed cells were then washed twice with PBS and mounted in a DAPI/Vectashied solution to stain the nuclei.

Fluorescence and bright-field images were captured using a Carl Zeiss LSM 510 Meta inverted confocal laser-scanning microscope, equipped with a 63 \times NA 1.4 planachromat objective. Z-series images were obtained from the bottom of the coverslip to the top of the cells. Each 8-bit TIFF image was transferred to the ImageJ software to quantify the total brightness and pixel area of each region of interest. For data analysis, the pixel areas of each cluster in the cytosol $s_i(\text{cyt})$, nucleus $s_i(\text{nuc})$ and lysosome $s_i(\text{lys})$ were separately summed in each XY plane. These values [$S'_{Z=j}(\text{cyt})$, $S'_{Z=j}(\text{nuc})$ and $S'_{Z=j}(\text{lys})$, respectively] were further summed through all of Z-series of images and denoted $S(\text{cyt})$, $S(\text{nuc})$ and $S(\text{lys})$, respectively. The total pixel area for the clusters of labeled pDNA in the cells, $S(\text{tot})$, was calculated as the sum of $S(\text{cyt})$, $S(\text{nuc})$ and $S(\text{lys})$. The fraction of pDNA in each compartment was calculated as $F(k) = S(k)/S(\text{tot})$, where k represents each subcellular compartment (nucleus, cytosol or lysosome).

RESULTS

Effect of iterative transfection on transgene expression

Previous work has led to the screening of human MARs to identify one, termed MAR 1–68, that was found to

potently increase and stabilize gene expression in cultured cells and in mice when inserted upstream of the promoter/enhancer sequences (21,39). Co-transfection of a GFP expression vector and an antibiotic resistance plasmid, followed by antibiotic selection of cells having stably integrated the transgenes in their genome, typically yields a bimodal distribution of the fluorescence in polyclonal cell populations when analyzed by flow cytometry (Figure 1A). A first cell subpopulation, which overlaps the y-axis in this experimental setting, corresponds to cells expressing GFP at undetectable levels, while another subpopulation of cells expresses significant GFP levels. Inclusion of MAR 1–68 increased the level of expression from fluorescent cells and concomitantly reduced the proportion of silent cells (15 versus 36%, Figure 1B). The increase in expression did not depend on the use of the viral SV40 promoter, as it was also obtained when expressing GFP from the the cellular glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene promoter (Supplementary Figure S1).

We next tested whether two consecutive co-transfections might further increase GFP expression from the MAR-containing plasmids (Figure 1A). When the same GFP expression vector was co-transfected again 2 weeks later with a distinct antibiotic resistance gene, a 2.4-fold increase of fluorescence was observed on average after selection for resistance to the second antibiotic, which is close to the expected 2-fold increase (Figure 1A and C). In contrast, an unexpectedly higher (4- to -5-fold) increase of GFP expression was observed from two successive transfections performed on consecutive days followed by selection with both antibiotics. When averaging over all cells of the polyclonal population, a 20-fold increase of expression was gained by successive transfections of MAR-containing plasmids when compared to a single transfection without a MAR (Figure 1C). Furthermore, some of the cells displayed very high levels of expression, and the occurrence of silent cells was almost fully abrogated from the polyclonal population (0.5%, Figure 1B). Consecutive transfections without a MAR yielded modest GFP expression, resulting in a 3.2-fold average increase of the overall fluorescence level when compared to a single transfection, and it did not abrogate the occurrence of silent cells (Figure 1C and data not shown). Thus, the presence of the MAR and the iterative transfection act synergistically to mediate elevated expression levels.

Overall, the expression levels obtained from the two consecutive transfections of MAR-containing plasmids were so high that the GFP fluorescence could be readily seen from the cell culture monolayers in the daylight, without excitation with UV light (Supplementary Figure S2A). This effect was not limited to the human MAR 1–68, as both the relatively less potent chicken lysozyme MAR (cLysMAR) and MAR X-29, a potent MAR isolated from human chromosome X (21), yielded an increase in expression when comparing double to single transfections (Supplementary Figure S3). This indicated that the elevated expression upon successive transfections may be a general property of MAR elements. However, the cLysMAR element yielded relatively lower transgene

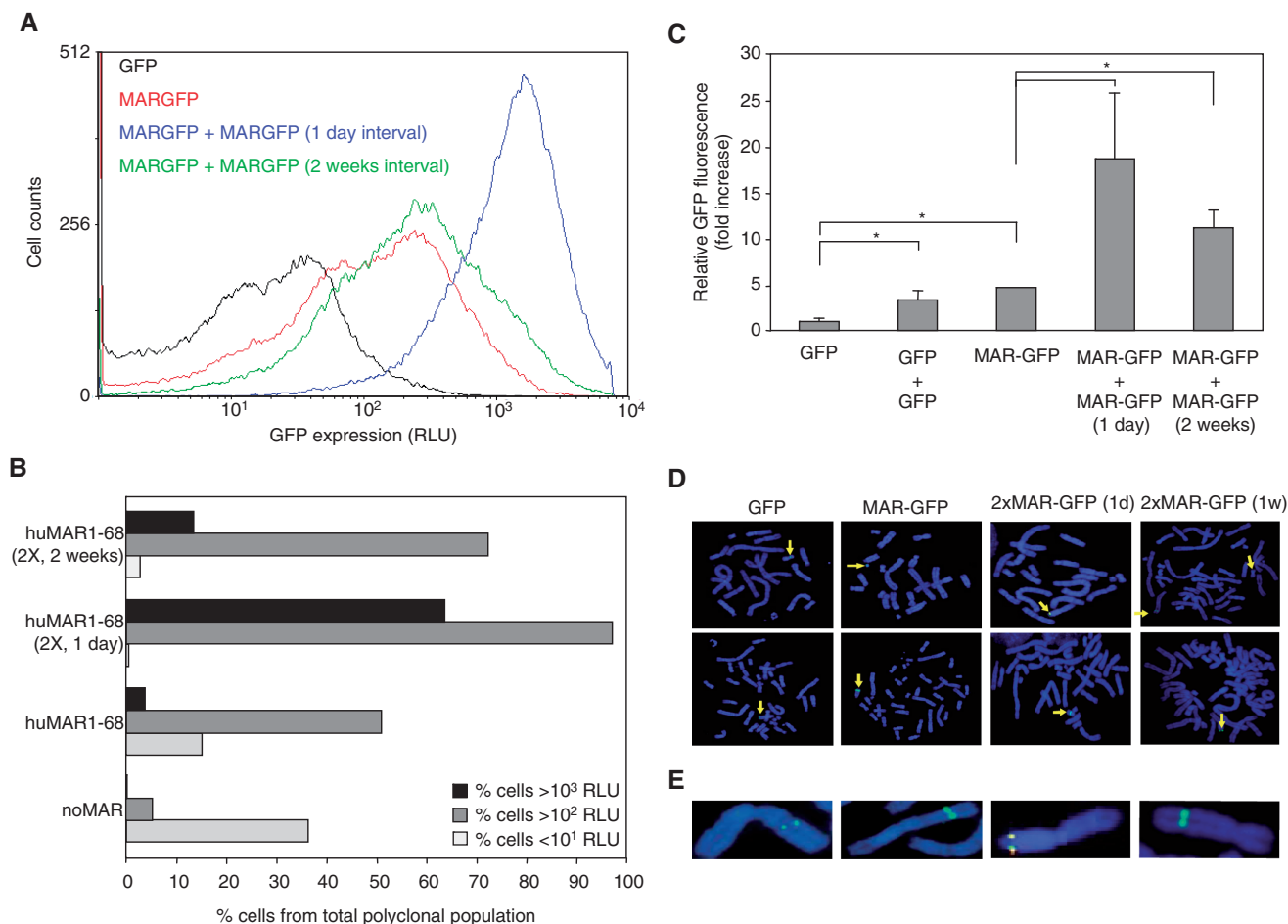


Figure 1. Analysis of the effect of MARs and successive stable transfections on gene transfer and expression. CHO DG44 cells were co-transfected with the GFP expression vector devoid of MAR element (GFP, dark line), or with the vector containing MAR 1–68 (MAR1–68GFP, red line), and with the pSVpuro plasmid mediating resistance to puromycin. Some of these cells were submitted to a second transfection with the same GFP expression vector but with a selection plasmid mediating neomycin resistance, either on the day following the first transfection (blue line) or after 2 weeks of selection for puromycin resistance (single transfections), or 3 weeks of selection for puromycin and neomycin resistance (double transfections at 1 day interval), or 2 weeks of selection for puromycin followed by the second transfection and two weeks of neomycin resistance selection (double transfections at 2 weeks interval), eGFP fluorescence was quantified by cytofluorometry. **(A)** Fluorescence distribution in polyclonal populations of GFP-expressing cells. The cell fluorescence profiles shown are representative of four independent experiments. **(B)** Histogram showing the percentage of total cells corresponding to non/low-expressors that display <10 relative light units (RLU), or cells that display medium and high (>100 RLU) or very-high (>1000 RLU) GFP fluorescence, as determined from the analysis of stable cell pools as shown in panel A. **(C)** The mean GFP fluorescence of each stable polyclonal cell pool was normalized to that obtained from the transfection of MARGFP and the average and standard deviation of four independent transfections is shown as a fold increase over the fluorescence obtained by one transfection without a MAR. Asterisks indicate significant differences in GFP expression (Student's *t*-test, $P < 0.05$). **(D)** FISH analysis of eGFP transgene chromosomal integration sites in cells singly or doubly transfected with or without the human MAR. Metaphase chromosome spreads of stable cell pools were hybridized with the GFP plasmid without MAR, and representative illustrations of the results are shown. **(E)** Enlargements of chromosomes are shown to illustrate differences in fluorescence intensities.

activation after a single transfection, and the increase in expression observed after the successive transfections was comparably lower. The converse was true of MAR X-29, indicating that MARs ability to mediate each of these effects may vary similarly.

The effect of the double transfection of MAR-containing plasmids was not dependent either on the GFP transgene or on the SV40 promoter used to express GFP, as similar results were obtained when a CMV promoter was used to express the DsRed reporter gene or the immunoglobulin light and heavy chains (Supplementary Figure S2B and data not shown). Interestingly, the very high levels of immunoglobulins

expressed by monoclonal CHO cell clones often correlated with an increased cell division time. This indicates that the cells were likely reaching their physiological limits in terms of protein synthesis. This may be expected, as cells synthesizing similar amounts of the recombinant protein as compared to their own cellular proteins (~100 pg per cell) should double the energetic input required for each cell division. Nevertheless, a large proportion of clones were found to express the heterologous protein at very high levels without interfering with their own metabolism, as they did not slow down cell division significantly (Supplementary Figure S2B).

Cointegration of transgenes upon iterative transfections

An important parameter driving high expression upon iterative transfection was found to be the time interval between the two transfections. The synergistic effect on expression was not observed when re-transfecting cells after 2 weeks. Rather, the two transfections behaved as two independent and thus additive events (Figure 1C). This suggested that the plasmid DNAs from each transfection may have to interact as episomes within the nucleus and potentially form mixed concatemers before co-integrating into the cell genome. This possibility was assessed by FISH analysis of metaphase chromosomal spreads from stable polyclonal populations. Eighty individual metaphases of cells transfected once either with or without the MAR element were hybridized with a probe consisting of the GFP plasmid without a MAR. A single integration site was observed, but higher fluorescence intensities were observed from cells transfected with the MAR (Figure 1D and E). Fluorescence intensity was further increased by the double transfection process, suggesting that a higher number of transgene copies had integrated. Unique integration sites were noted in all cases after single or two consecutive transfections. However, double integration events were observed in approximately half of the cells transfected twice at an interval of 1 week, when little episomal DNA should remain from the first transfection. This indicates that independent integration events may occur if DNA integration from the first transfection has been completed before the second transfection is performed. Double transfections did not lead to apparently increased aneuploidies nor to detectable chromosomal rearrangements, and they did not detectably lead to insertions at a preferred chromosomal locus, as none of the analyzed cells had an identical integration site. Thus, transgene integration upon two transfections does not appear to be targeted to any specific chromosomes or chromosomal sites, as reported earlier for single transfections of MAR-containing plasmids (21).

High transgene expression requires phasing of the cell cycle and transfections

As timing between transfections seemed to play a role in high transgene expression, we analyzed the effect of systematic variations of the time interval. The highest GFP expression level was observed when the second transfection was performed 21 h after the first one, yielding consistently a 5-fold increase of fluorescence as compared to a single transfection. When the second transfection was performed after 18, 24 and 27 h, a 3- to 3.5-fold increase of expression was obtained as compared to a single transfection, but these were significantly lower than those obtained after 21 h (Figure 2A). As this timing is close to the duration of the first cell division cycle after cell passaging (Supplementary Figure S4), this suggested that high transgene expression upon consecutive transfections might be linked to particular phases of the cell division cycle.

The distribution of the cells along the division cycle was determined by PI staining of the DNA. This analysis

indicated an over-representation of cells at the G1 phase 18 h after cell passaging, and this was found to correspond to the timing that yields the highest expression from a single transfection (Figure 2B, Supplementary Figure S4A and B and data not shown). A similar pattern and over-representation of G1 cells was obtained 21 h after the first transfection, which again corresponds to the timing that yields the highest expression levels upon a second transfection (Figure 2B). If expression is indeed linked to cell cycle phasing, another optimum for transgene expression should be observed if a second transfection was performed at an interval corresponding to two cell divisions. After 42 h, the synergistic effect of the two transfections was lost, as expression was similar to that obtained for one transfection. However, a second, albeit lower, synergistic increase of transgene expression was observed after 48 h. The higher levels of expression observed from a first transfection at 18 h and for a second transfection performed with a 21 or 48 h interval imply that optimal DNA transfer and/or expression may occur at specific cell division stages.

Effect of MAR and consecutive transfections on cellular DNA uptake

FISH analysis suggested that elevated expression upon successive transfections may result in part from the integration of a higher number of the transgene copies in the genome (Figure 1D). Consecutive transfections at an interval of one day might lead to an increase of the concentration of plasmid episomes in the nucleus, thereby augmenting the probability of transgene integration within the cell genome. To assess the amount of transgene entering nuclei at each transfection, we performed transient single or double transfections followed by plasmid extraction from nuclei isolated 1 or 2 days after the second transfection and quantification of the transgenes by real-time quantitative PCR (qPCR). Cells doubly transfected with MAR-GFP exhibited 3.8-fold more GFP transgene copies in their nuclei than cells transfected just once with MAR-GFP (Figure 3A). When comparing cells transfected with these different plasmids expressing either GFP or DsRed, we observed that the nuclear delivery resulting from the second transfection of MAR-GFP was 4.2-fold higher than the one observed from a single transfection of this plasmid. However, the nuclear transport of the firstly transfected GFP plasmid was not increased significantly by performing a second transfection. We concluded that DNA transport to the nucleus from the second transfection is favored by performing a prior first transfection.

These conclusions were strengthened by confocal imaging of DNA transport, where plasmids used for the first transfection were labeled with rhodamine while the secondly transfected plasmids were labeled with Cy5 (red and white labels, respectively, Figure 4A). Similar numbers of rhodamine-labeled plasmid clusters were observed in cell nuclei after a first transfection with or without a MAR, which correlates well with the lack of effect of the MAR on DNA transport as assessed by qPCR (Figures 3A and 4A). Nuclear plasmid clusters

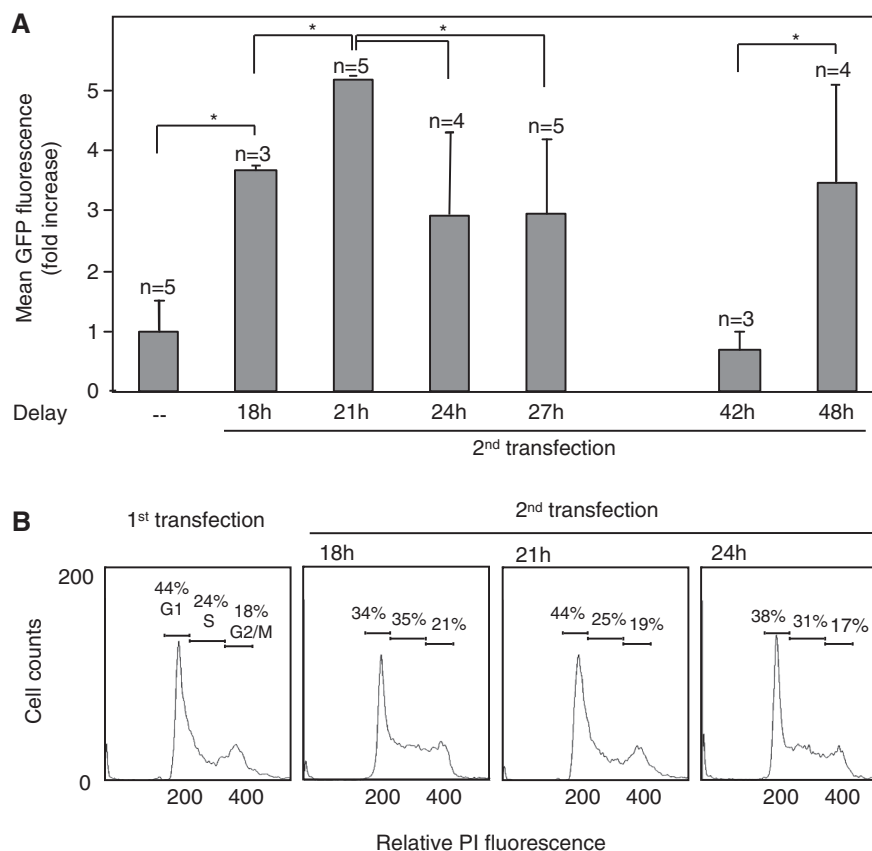


Figure 2. Determination of the optimal timing between successive transfections. **(A)** Stable polyclonal populations were generated by a single transfection (minus symbol) or by two consecutive transfections of the MAR-GFP expression plasmid with the indicated time intervals. After 2 weeks of selection, mean GFP expression of the total polyclonal populations was determined. Fluorescence levels were normalized to the maximal values obtained and they are displayed as the fold increase over the expression obtained from a single transfection where n corresponds to the number of independent transfections. Asterisks indicate significant differences in GFP expression (Student's t -test, $P < 0.05$). **(B)** Analysis of cell cycle progression. At the time of first and second transfections, CHO cells were harvested and stained with propidium iodide (PI) and fluorescence was analyzed by cytofluorometry. The distribution of relative PI fluorescence represents the amount of genomic DNA per cell. The percentage of the population associated to each cell cycle state (G1, S, G2/M) is as indicated.

were observed in essentially all the cells after two transfections. However only few cells expressed GFP, in agreement with previous observations that only a minority of cells are able to express transiently transfected genes (58).

The transport of transfected plasmid DNA in CHO cells, which is known to consist of cellular uptake, lysosomal escape and nuclear import, is limited by endosomal/lysosomal degradation (58). Thus, we next assessed the intracellular trafficking of transfected plasmid DNA by quantifying its distribution in cellular organelles and in the cytosol after each transfection, after specific staining of the endosomal/lysosomal and nuclear compartments to distinguish them from the cytosol. Results summarized in Figure 4B show a similar subcellular distribution of plasmid DNA with or without MAR 21 h after a first transfection, although nuclear transport of MAR-containing plasmids seemed somewhat faster at the earlier time points. Performing a second transfection of the MAR-devoid plasmid did not yield an improved nuclear transport. However, plasmids bearing a MAR element escaped lysosomal retention and entered nuclei much more efficiently, as 80% of the total Cy5-labeled pDNA was located in the nuclei in presence of the

MAR 21 h after the second transfection, as compared to <40% of the plasmid devoid of MAR (Figure 4B). Rather, most of the MAR-devoid plasmid ended up in the lysosomal/endosomal compartment, as found also for the first transfection. The unexpected finding of a cooperative effect of the MAR and of the iterative gene transfer on lysosomal escape thus provides an explanation for the increased concentration of plasmids in isolated nuclei (Figures 3A and 4B). The reason for this phenomenon is unclear at present, but it might result in part from the saturation of the cellular degradation compartments by the DNA of first transfection, thus allowing plasmids of the second transfection to remain in the cytoplasm where the MAR may promote plasmid transport into the nucleus.

MAR elements increase the copy number of genome-integrated transgenes

Next, we examined whether the increased transport of plasmid DNA elicited by the MAR and the consecutive transfections may increase transgene integration into the genome of CHO cells. Stable polyclonal cell populations

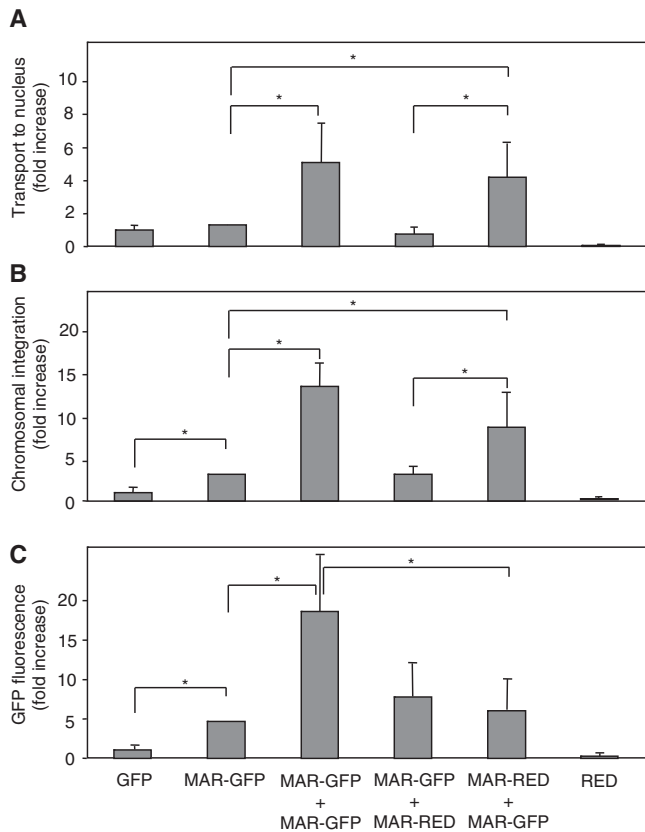


Figure 3. DNA transport, integration and expression upon successive transfections. (A) Amount of GFP transgenes transport into cell nuclei during single and double transient transfections with GFP or DsRed plasmids with or without a MAR. MAR-GFP + MAR-RED corresponds to a double transfection where MAR-GFP is transferred during the first transfection, whereas MAR-RED was used in the second transfection. Nuclei were isolated and total DNA was extracted one day after a single or after the second transfection, and the number of GFP transgenes transported into the nuclei was quantified by qPCR. Results were normalized to that of the reference CHO cell genomic GAPDH gene and represent the mean of four independent transfections. (B) Effect of the MAR and successive transfections on integrated GFP transgene copy number. Total genome-integrated transgene DNA was extracted from the previously described GFP-expressing cells after 3 weeks of selection of stable polyclonal cell pools, and DNA was quantified as for panel A. (C) Effect of MAR and successive transfections on GFP expression. The GFP fluorescence levels of the stable cell pools analyzed in panel B were assayed by cytofluorometry.

were selected as for Figure 1, and the average numbers of stably integrated GFP transgene copies per genome were determined on total cell DNA using qPCR. Inclusion of a MAR element in transfected plasmids significantly increased the number of transgenes integrating in the genome of stable cell pools (Figure 3B). Because the MAR does not act to increase nuclear transport after single transfections (Figure 3A), we concluded that the MAR may increase genomic integration of the plasmid *per se*. This finding supports previous indications that the use of MARs may increase the number of transgene copies that integrate in the genome of recipient cells (51,59).

Successive transfections also mediated a 4-fold increase of plasmid integration, which is commensurate to the

increase in free episomes noted in transient transfections (Figure 3A and B). We estimated that 48 GFP plasmid copies had integrated on average when transfecting once without a MAR, while ~163 copies and 676 copies on average were obtained from one or two successive transfections with the MAR, respectively. Overall, the increased nuclear transport synergistically elicited by both the MAR and the successive transfections yielded a >10-fold increase in transgene copy number when combined to the MAR-driven increase of plasmid integration. Note, however, that the double transfection of MAR-containing plasmids yielded yet an even higher increase in transgene expression (~20-fold, Figure 3C). This indicated that increased expression did not result solely from increased transgene copy number, but that it must also stem from a MAR-mediated increase of expression per integrated transgene copy, as expected from the previously observed antisilencing and transcription activation effects of this MAR element (39).

When assessing GFP expression and transgene copy number in individual cell clones isolated from the polyclonal populations, a correlation was found between transgene expression and copy number, in that higher levels of integration and of expression were observed from single or multiple transfections of MAR-containing plasmids (Supplementary Figure S5A). Furthermore, no significant decrease of expression could be detected from MAR-containing clones having co-integrated very high numbers of transgene copies and MARs (Supplementary Figure S5B), and stable elevated expression was maintained upon adaptation of several of these clones to growth in suspension and further culturing (data not shown). These results indicated that the MAR was able to prevent inhibitory effects that may result from the repetitive nature of the co-integrated plasmids and/or from antisense transcription, an effect that can be attributed to the potent anti-silencing properties of this MAR element (39). However, the average levels of expression did not match perfectly with copy number, as noted when analyzing individual cell clones (Supplementary Figure S5A), or when comparing GFP expression from the firstly or secondly transfected DNA, in co-transfection experiments with the dsRED vector (Figure 3B and C). We therefore conclude that the enhanced transgene expression observed after two successive transfections of MAR-containing plasmids can be explained in part by the improved nuclear import and genomic integration of the plasmid DNA, which results in increased transgene copy number, as well as by the lack of silencing and higher transgene transcription. However, other effects may also influence transgene expression in individual clones depending on the transfection history and conditions, clonal fitness heterogeneity and/or possible effects pertaining to the genomic integration locus.

Effects of DNA homology on plasmid integration and expression

As the high GFP fluorescence observed from successive transfections of MAR-containing plasmids results in part from the increased transgene integration at a single

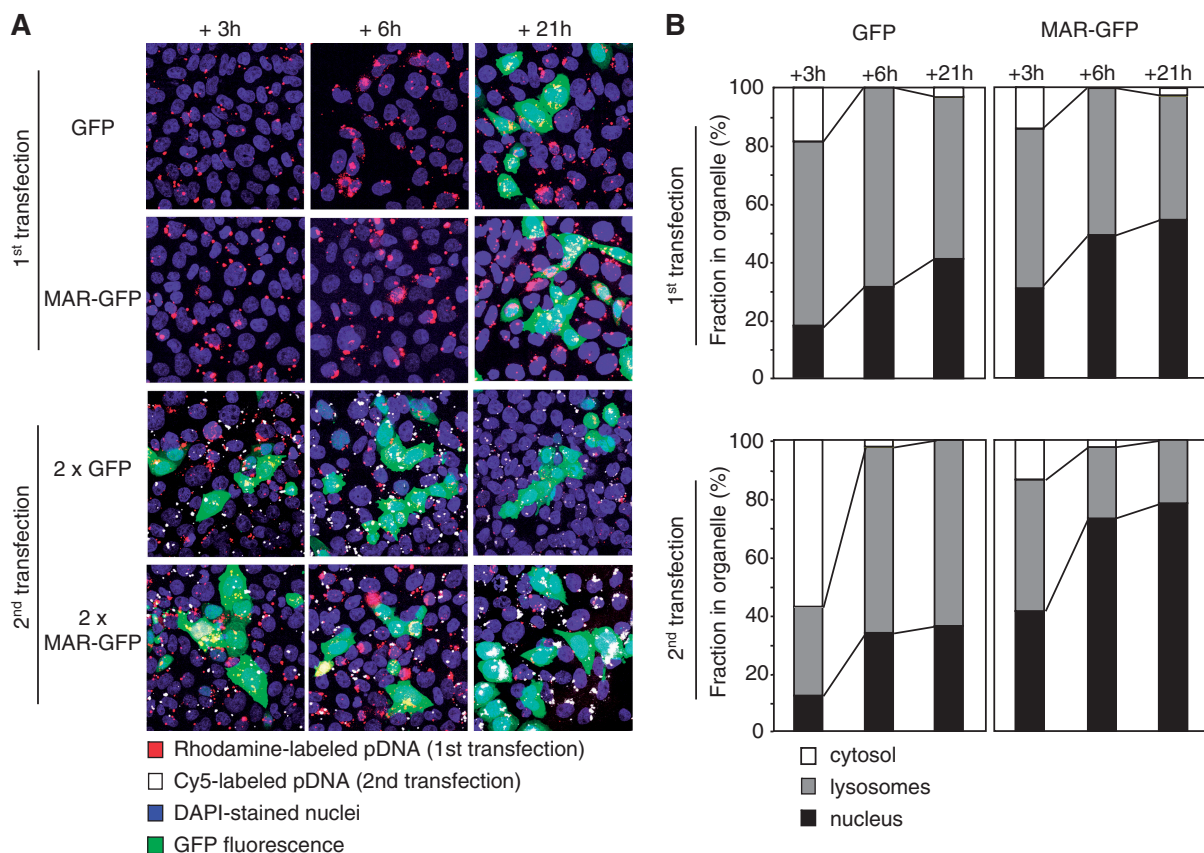


Figure 4. Subcellular distribution of transfected DNA. (A) Confocal microscopy analysis of DNA intracellular trafficking. Transient single or double transfections were performed in CHO cells using plasmids bearing or not a MAR labeled with Rhodamine and Cy5 fluorophores, as indicated. Transfected cells were fixed and stained with DAPI (blue) 3, 6, 21 h post-transfections. Cells expressing GFP appear in green on the pictures. (B) Quantification of the subcellular plasmid DNA distribution was performed on confocal laser microscopy performed for panel A, except that endosome/lysosome compartments were stained with LysoTracker Red DND-99. The pixel area of clusters derived from rhodamine or Cy5 fluorescence were used to estimate the amount of plasmid DNA in ~120 cells.

chromosomal locus, we next examined the molecular basis of this effect. A possibility is that the integration of a MAR-containing plasmid during the first transfection might promote secondary integration at the same genomic locus during the second transfection. This may be expected from the ability of the MAR to maintain chromatin in an accessible state and thus to provide proper targets for HR. Alternatively, the high number of integrated transgenes may result from a more efficient concatemerization of the plasmids introduced during both transfections, as may be mediated by the high concentration of episomes found in the nucleus. Indeed, HR was proposed to mediate the formation of large concatemers of transfected plasmids (60), which may lead to the co-integration of multiple plasmid copies upon recombination with the genomic DNA. In the latter model, HR may occur between similar DNA sequences on the plasmids used during the first and second transfections, and thus the efficacy of transgene integration and expression should critically depend on DNA sequence homologies.

This latter possibility was first assessed by analyzing the effect of plasmid homology on transgene expression by performing successive transfections with different

combinations of transgenes (GFP or DsRed), plasmid backbones (ampicillin or kanamycin bacterial resistance) and/or MARs (chicken lysozyme MAR or the human MAR 1–68). Transfection of distinct MARs, transgenes, or bacterial resistance all decreased the high expression normally observed with successive transfections (Figure 5A). The double transfection effect was almost fully abolished when using different MARs, transgenes and vector elements (MAR₁-GFP + MAR₂-RED constructs), suggesting that plasmid homology is required to achieve high expression from successive transfections.

HR is often elicited as a DNA repair mechanism of double-stranded breaks (DSB), in a process that was termed homologous recombination repair (61,62). Thus, we tested whether plasmid linearization prior to transfection mediates the high expression obtained from successive transfections. A more than additive increase of transgene expression was also observed with circular plasmids, however, the overall expression was lower than that obtained using linear plasmids (Supplementary Figure S6). This is consistent with the increased recombinogenic properties of linear DNA in HR processes (63).

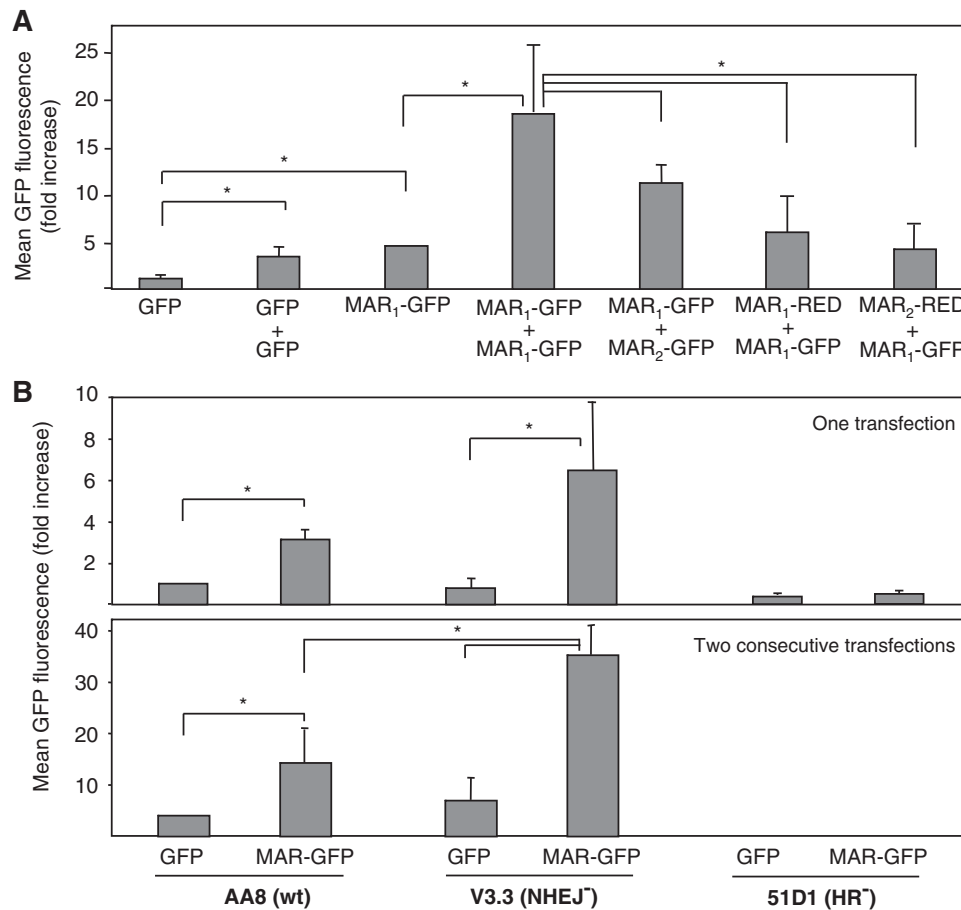


Figure 5. The MAR, plasmid homology and homologous recombination mediate high transgene expression. (A) Stable polyclonal cell pools were generated by the transfection of plasmids bearing different transgenes (GFP or DsRed), MAR (MAR1 for the human 1–68 or MAR2 for the chicken lysozyme MAR), and/or bacterial resistance gene (ampicillin or kanamycin), and the relative average GFP fluorescence of four independent transfections are shown as the fold increase over that obtained from one transfection without MAR. Asterisks show significant differences in GFP expression (Student's *t*-test, $P < 0.05$). (B) Stable transfections with GFP or MAR1-68GFP plasmids were performed in the parental CHO cell line (AA8) and in mutants deficient either in the homologous recombination (51D1) or non-homologous end-joining (V3.3) pathway. The mean GFP fluorescence of each stable polyclonal cell pool generated from single (top panel) or two consecutive (bottom panel) transfections were normalized to that obtained from AA8 cells singly transfected with the MAR-devoid plasmid. Asterisks indicate significant differences in GFP expression (Student's *t*-test, $P < 0.05$). No stably transfected cells were obtained from the double transfection of 51D1 cells.

HR mediates increased expression

The requirement of plasmid homology and double-strand breaks to achieve the higher expression levels upon the double transfection implied that HR may be involved. Transgenes were proposed to integrate into the cell genome using two families of antagonistic pathways, termed non-homologous end-joining (NHEJ) or HR. These pathways are more active during specific phases of the cell cycle, as exemplified by HR, which is used to repair DNA damages during or after DNA replication in the S and G2/M phases of the cell cycle (64). Cells lacking classical NHEJ genes show a double-stranded break repair biased in favor of HR, suggesting that these two major pathways normally compete to repair these DNA lesions (65). Thus, one way to activate HR is to suppress or genetically inactivate NHEJ, as seen in yeast and mammalian cells (65–68). A possible implication of HR-related mechanisms in the increased transgene expression that results from the MAR and/or successive transfections was thus directly

assessed using CHO cell lines mutated in a key component of either pathways, and which are thus only competent for either HR or NHEJ.

The 51D1 CHO mutant derivative lacks the RAD51 strand transferase and is thus deficient in HR, while V3.3 CHO cells lack the catalytic activity of the DNA-dependent protein kinase (DNA-PK) that plays an essential role to initiate the NHEJ pathway (13,54,69). In the wild-type parental cell lines (AA8), the MAR mediated a 3-fold increase of the overall GFP fluorescence in the polyclonal population (Figure 5B). However, few stably transfected colonies survived after selection for antibiotic resistance in the 51D1 cell line and GFP expression remained very low. In contrast, an exacerbated MAR-driven activation of transgene expression was observed in NHEJ-deficient cells, resulting in a >6-fold increase of transgene expression when compared to cells transfected once with the GFP expression vector without MAR.

Similar trends were noted for successive transfections, in that GFP expression from V3.3 cells was increased in

the presence of the MAR as compared to the parental AA8 cells (Figure 5B, note the different scales of the top and bottom panels). Inactivation of the NHEJ pathway had little effect on the expression of the MAR-devoid plasmid but it further exacerbated the expression increase elicited by the MAR, indicating that presence of the MAR and high HR activity are both necessary to obtain very high transgene expression. Cells deficient in HR did not yield stable colonies from the double transfection, demonstrating the requirement of the HR pathway in the efficient integration and maintenance of transgenes in the successive gene transfer process. Analysis of the number of integrated transgene copies revealed a 25-fold higher number of integrated transgene copies upon the double transfection of MARGFP in NHEJ-deficient V3.3 cells (Supplementary Figure S7). However, a relatively larger (over 35-fold, Figure 5B) increase in expression was obtained from two consecutive transfections of the V3.3 cells with the MAR. Again, this indicated that the observed MAR-mediated increase in expression results both from an increase in transgene genomic integration and in an augmentation of the expression per transgene copy.

DISCUSSION

The variability in gene expression among independently transformed cells or organisms is well documented. Inconsistent expression levels have been associated to the variable number of genes that have stably integrate in the host cell genome, to properties of the sites of integration, and/or to the gene transfer procedures. Non-viral gene transfer remains characterized by variable expression efficiencies and by an uncertain outcome in terms of expression levels and stabilities. Gene amplification has been used to augment the copy number and hence expression of the exogenous genes (70), however this often leads to unstable expression when the selection reagent is removed (71). An alternative has been to optimize expression vectors by inserting synthetic or natural regulatory sequences that increase and/or stabilize expression, such as MAR elements (39). However, even in the most favorable conditions, transfections have lead to the occurrence of cells that integrate the transgene in their genome but express it at low levels. Hence, the identification of stable cell lines is usually associated with the tedious isolation and characterization of many clones to identify one with the desired expression properties. Here, we show that very high levels of transgene expression can be consistently obtained in nearly all cells of a polyclonal cell population after successive transfections of MAR-containing constructs. Our results further indicate that efficient gene transfer and expression of MAR-driven vectors require a functional HR or repair pathway.

In this study, we show that MAR elements can act in part to increase the number of copies of exogenous genes that integrate in the genome, substantiating previous non-quantitative observations (21,59). However, the intuition that a high copy number always supports stronger expression is often non-valid, as the presence of multiple

gene copies co-integrated at one or a few loci of the host genome has been reported to favor silencing. The propensity of repeated elements to pair and assemble in heterochromatin or to generate double-stranded and/or small interfering RNAs from antisens transcription of adjacent transgenes is a frequent cause of gene silencing (6,72). Here, the copy number of integrated transgenes and cell fluorescence levels were shown to correlate well when comparing single or double transfections performed in the presence of the MAR. We find also that the MAR-mediated increase in transgene expression results in part from the integration of more transgene copies in the host cell genome. However, we also find that the relative increase of transgene expression is higher than the increase of transgene copy number upon inclusion of the MAR. This indicates that MAR 1-68 acts both to increase transgene integration and to increase expression per transgene copy. Overall, we estimate that increased integration mediates ~70% of the observed effect on the increase of expression, while the augmentation of transgene transcription per se, as shown to occur in presence of MAR 1-68 (39), accounts for the remaining 30% of the increase of expression observed upon successive transfections. Thus, both effects concur to mediate the very high expression levels observed after successive transfections of MAR-containing plasmids.

We also find that transgene genomic integration is synergistically increased by successive gene transfers and by the MAR. Successive transfections of MAR-containing plasmids result in improved plasmid transport to the nucleus, as explained by the decreased targeting of the second DNA load to the degradation compartment. While the greater amounts of episomal plasmids in the nucleus may readily explain the increased co-integration of independent DNA loads in successive transfections (Supplementary Figure S8), we find that the MAR-mediated increase in transgene integration does not result from an increased plasmid targeting or transport to the nucleus. This indicated that the MAR may directly promote the recombination of the exogenous DNA with the host cell genome.

In vertebrates, HR and NHEJ differentially contribute to repairing abnormal DNA structures such as double-stranded breaks, depending on the nature of the DNA damage and the phase of the cell cycle (64). This study shows that plasmid integration critically depend on a functional HR pathway, as inactivation of Rad51, a key initiator of HR (73), nearly abolishes transgene integration and expression. Conversely, inactivation of the antagonistic NHEJ pathway, which is associated to increased HR, concomitantly increases the integration and expression of MAR-containing plasmids, but not of the MAR-devoid control. In addition, increased integration and expression upon successive MAR transfections require the plasmids to have homologous DNA sequences. Furthermore, the cells must be in the G1 cell cycle phase at the time of transfection, and we find that DNA transport to the nucleus is nearly completed within 6h after DNA transfer, at which time cells have progressed to exit S and enter G2 (see Supplementary Data), at which time HR is most active. Taken together, these findings provide strong

support for the involvement of HR in the increased integration, and, consequently, the higher expression observed in presence of the MAR and/or upon double transfections. They also suggest that plasmid integration by a HR repair pathway may be increased in the presence of a MAR element on the DNA. However, a direct demonstration of the role of MAR elements in activating HR will require further experimentations.

Mammalian cells contain the enzymatic machinery required to mediate recombination between newly introduced plasmid DNA molecules, and HR between co-injected plasmid molecules is an efficient process in cultured mammalian cell lines, approaching 100% of the molecules (60). Plasmid concatemers thereby formed may integrate at one or a few sites in the host chromosome, the integration site being different in independently transformed cells (74). The orientation of the copies within the concatemer is not random, but usually organized as tandem head-to-tail arrays, as generated by the HR of independent plasmid copies in the cell (75). HR between the newly introduced DNA and its homologous chromosomal sequence has been reported to occur infrequently in the absence of MAR element, at a frequency of 1:1000 cells receiving DNA (76). However, estimations based on transgene expression, such as antibiotic selection, may significantly underestimate the true frequency of integration. Whether plasmid concatemers may be able to undergo HR with previously integrated transgene copies, or whether a single integration event of a large plasmid concatemer occurs, remains difficult to assess experimentally. In any case, the contribution of MAR elements to promote HR proposed in this study readily provides an explanation to the previously observed MAR-mediated increase in the occurrence of cells that stably integrate the transfected genes, and thus in the number of antibiotic-resistant colonies (21).

The requirement of a functional HR pathway to mediate efficient integration of MAR-containing constructs implies that these DNA elements might be preferred sites of homologous pairing and DNA recombination. A number of molecular mechanisms have been proposed to explain the mode of action of MARs. For instance, they may reduce the occurrence of lowly-expressing cells by protecting transgenes from silencing effects linked to integration at heterochromatic loci (7). MARs may recruit regulatory proteins that modify chromatin to adopt a more transcriptionally permissive state by mediating histone hyperacetylation, they may change subnuclear localization of the transgene, or they may facilitate the association of transcription factors that activate gene transcription (2,8,36,39). Although not mutually exclusive, our results indicate that a significant part of the effect of MAR elements on transgene expression might in fact be attributed to increased transgene integration into the host cell genome by HR.

The proposed role of MARs as HR-promoting genomic signals might result from their action on chromatin and/or on DNA accessibility, thereby providing an access to DNA binding proteins such as topoisomerases (48). As reported in this study and by prior reports, all cells take-up the transgene during a transfection although few

express it at a detectable level. This effect remains poorly understood, but it may result from the slow or inefficient unpacking of the DNA from complexes generated by the transfection reagents (77). Thus, the MAR might potentially act in part by facilitating the release of the DNA from complexes with the transfection agent and/or with repressive protein structures. However, specific DNA structures may also act to promote recombination. For instance, endonuclease-mediated double-stranded DNA breaks mediate improved rates of HR in eukaryotes, bacteria and archae (78–82). Thus, an involvement of the unwinding and DNA strand unpairing potential of MARs might directly or indirectly promote the association of recombination-initiating proteins (22,23,83,84). Finally, MAR elements have been associated with the promotion of DNA replication of episomally maintained vectors (28,29). However, this latter effect requires a promoter mediating transcription of the MAR, which is not the case in this study. Furthermore, the MAR sequence used in this study did not lead to increased episomal DNA, and transgene maintenance was clearly associated to chromosomal integration. Thus, if a direct role for DNA replication appears unlikely in the settings used in this study, the requirement of proteins such as replication promoting activity in both replication and HR would be consistent with a role of MARs in promoting both activities.

Overall, our findings imply that successive transfections and the MAR may mediate very efficient expression by promoting HR between individual plasmid molecules, thereby favoring the chromosomal integration of larger concatemers, and by maintaining a permissive chromatin structure after genomic integration. An interesting but as yet unexplored possibility might be that the MAR and HR-mediated events may lead to transgene integration at regions of homologies with the cellular genome. One tantalizing possibility might even be that genomic integration might occur at the endogenous cellular MAR elements, as may result from the fact that MARs share similar AT-rich and highly repetitive sequences across species. Thus, the transgenes would be expected to integrate at more accessible or privileged regions of the genome, explaining the efficient and stable expression. It will be interesting to evaluate whether MAR elements may contribute to regulating HR in a chromosomal context and whether this can be exploited to facilitate gene replacement strategies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Arnaud Paradis for skilful help in confocal microscopy and imaging, and Dr Fabrizio Palitti for the kind gift of materials. The authors thank Nathalie Besuchet-Schmutz for excellent technical assistance.

FUNDING

This project was funded by a grant from the Swiss Commission for Technology and Innovation and Selexis SA, and by the University of Lausanne. Funding for open access charge: University of Lausanne.

Conflict of interest statement. Some genetic elements used in this study are also being used by a biotech SME, Selexis SA, co-founded by N.M. After the completion of their work, the company hired some co-authors (P.A.G., D.C.) to work in its own laboratories. This work was financed jointly by a Swiss government agency and by the University, and for a smaller part by the company. We are confident that this has not affected our objectivity in presenting these results, as these were obtained and written-up independently of the authors who joined the company.

REFERENCES

- Kalos, M. and Fournier, R.E. (1995) Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. *Mol. Cell. Biol.*, **15**, 198–207.
- Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci USA*, **99**, 6883–6888.
- Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995) Position-dependent variegation of globin transgene expression in mice. *Proc. Natl Acad. Sci. USA*, **92**, 5371–5375.
- Henikoff, S. (1996) Dosage-dependent modification of position-effect variegation in *Drosophila*. *Bioessays*, **18**, 401–409.
- Wakimoto, B.T. (1998) Beyond the nucleosome: epigenetic aspects of position-effect variegation in *Drosophila*. *Cell*, **93**, 321–324.
- Martienssen, R.A. (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat. Genet.*, **35**, 213–214.
- Bell, A.C. and Felsenfeld, G. (1999) Stopped at the border: boundaries and insulators. *Curr. Opin. Genet. Dev.*, **9**, 191–198.
- Hart, C.M. and Laemmli, U.K. (1998) Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.*, **8**, 519–525.
- Bode, J., Goetze, S., Heng, H., Krawetz, S.A. and Benham, C. (2003) From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Res.*, **11**, 435–445.
- Bode, J., Benham, C., Knopp, A. and Mielke, C. (2000) Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit. Rev. Eukaryot. Gene Expr.*, **10**, 73–90.
- Gasser, S.M. and Laemmli, U.K. (1987) Improved methods for the isolation of individual and clustered mitotic chromosomes. *Exp. Cell Res.*, **173**, 85–98.
- Bode, J., Schlake, T., Rios-Ramirez, M., Mielke, C., Stengert, M., Kay, V. and Klehr-Wirth, D. (1995) Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. *Int. Rev. Cytol.*, **162A**, 389–454.
- Jackson, D.A. (1997) Chromatin domains and nuclear compartments: establishing sites of gene expression in eukaryotic nuclei. *Mol. Biol. Rep.*, **24**, 209–220.
- Stein, G.S., van Wijnen, A.J., Stein, J.L. and Lian, J.B. (1999) Interrelationships of transcriptional machinery with nuclear architecture. *Crit. Rev. Eukaryot. Gene Expr.*, **9**, 183–190.
- Heng, H.H., Krawetz, S.A., Lu, W., Bremer, S., Liu, G. and Ye, C.J. (2001) Re-defining the chromatin loop domain. *Cytogenet. Cell Genet.*, **93**, 155–161.
- Heng, H.H., Goetze, S., Ye, C.J., Liu, G., Stevens, J.B., Bremer, S.W., Wykes, S.M., Bode, J. and Krawetz, S.A. (2004) Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J. Cell Sci.*, **117**, 999–1008.
- Mirkovitch, J., Mirault, M.E. and Laemmli, U.K. (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, **39**, 223–232.
- Capelson, M. and Corces, V.G. (2004) Boundary elements and nuclear organization. *Biol. Cell*, **96**, 617–629.
- Boulikas, T. (1993) Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. *J. Cell Biochem.*, **52**, 14–22.
- Homburger, H.P. (1989) Bent DNA is a structural feature of scaffold-attached regions in *Drosophila melanogaster* interphase nuclei. *Chromosoma*, **98**, 99–104.
- Girod, P.A., Nguyen, D.Q., Calabrese, D., Puttini, S., Grandjean, M., Martinet, D., Regamey, A., Saugy, D., Beckmann, J.S., Bucher, P. et al. (2007) Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat. Methods*, **4**, 747–753.
- Platts, A.E., Quayle, A.K. and Krawetz, S.A. (2006) In-silico prediction and observations of nuclear matrix attachment. *Cell Mol. Biol. Lett.*, **11**, 191–213.
- Evans, K., Ott, S., Hansen, A., Koentges, G. and Wernisch, L. (2007) A comparative study of S/MAR prediction tools. *BMC Bioinformatics*, **8**, 71.
- Ottaviani, D., Lever, E., Takousis, P. and Sheer, D. (2008) Anchoring the genome. *Genome Biol.*, **9**, 201.
- Koina, E. and Piper, A. (2005) An inactive X specific replication origin associated with a matrix attachment region in the human X linked HPRT gene. *J. Cell Biochem.*, **95**, 391–402.
- Mearini, G., Chichiarelli, S., Zampieri, M., Masciarelli, S., D'Erme, M., Ferraro, A. and Mattia, E. (2003) Interaction of EBV latent origin of replication with the nuclear matrix: identification of S/MAR sequences and protein components. *FEBS Lett.*, **547**, 119–124.
- Amati, B.B. and Gasser, S.M. (1988) Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold. *Cell*, **54**, 967–978.
- Piechaczek, C., Fetzter, C., Baiker, A., Bode, J. and Lipps, H.J. (1999) A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. *Nucleic Acids Res.*, **27**, 426–428.
- Jenke, A.C., Stehle, I.M., Herrmann, F., Eisenberger, T., Baiker, A., Bode, J., Fackelmayer, F.O. and Lipps, H.J. (2004) Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *Proc. Natl Acad. Sci. USA*, **101**, 11322–11327.
- Liebich, I., Bode, J., Reuter, I. and Wingender, E. (2002) Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Res.*, **30**, 3433–3442.
- Pederson, T. (1998) Thinking about a nuclear matrix. *J. Mol. Biol.*, **277**, 147–159.
- Cai, S., Han, H.J. and Kohwi-Shigematsu, T. (2003) Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat. Genet.*, **34**, 42–51.
- Feister, H.A., Torrungruang, K., Thunyakitpisal, P., Parker, G.E., Rhodes, S.J. and Bidwell, J.P. (2000) NP/NMP4 transcription factors have distinct osteoblast nuclear matrix subdomains. *J. Cell Biochem.*, **79**, 506–517.
- Dunn, K.L., Zhao, H. and Davie, J.R. (2003) The insulator binding protein CTCF associates with the nuclear matrix. *Exp. Cell Res.*, **288**, 218–223.
- Fernandez, L.A., Winkler, M. and Grosschedl, R. (2001) Matrix attachment region-dependent function of the immunoglobulin mu enhancer involves histone acetylation at a distance without changes in enhancer occupancy. *Mol. Cell. Biol.*, **21**, 196–208.
- Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P. and Kohwi-Shigematsu, T. (2002) SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature*, **419**, 641–645.
- Donev, R., Horton, R., Beck, S., Doneva, T., Vatcheva, R., Bowen, W.R. and Sheer, D. (2003) Recruitment of heterogeneous nuclear ribonucleoprotein A1 in vivo to the LMP/TAP region of the major histocompatibility complex. *J. Biol. Chem.*, **278**, 5214–5226.

38. Rajaiya, J., Nixon, J.C., Ayers, N., Desgranges, Z.P., Roy, A.L. and Webb, C.F. (2006) Induction of immunoglobulin heavy-chain transcription through the transcription factor Bright requires TFII-I. *Mol. Cell. Biol.*, **26**, 4758–4768.
39. Galbete, J.L., Buceta, M. and Mermod, N. (2009) MAR elements regulate the probability of epigenetic switching between active and inactive gene expression. *Mol. Biosyst.*, **5**, 143–150.
40. Kwaks, T.H. and Otte, A.P. (2006) Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends Biotechnol.*, **24**, 137–142.
41. Allen, G.C., Spiker, S. and Thompson, W.F. (2000) Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Mol. Biol.*, **43**, 361–376.
42. Namciu, S.J. and Fournier, R.E. (2004) Human matrix attachment regions are necessary for the establishment but not the maintenance of transgene insulation in *Drosophila melanogaster*. *Mol. Cell. Biol.*, **24**, 10236–10245.
43. Dang, Q., Auten, J. and Plavec, I. (2000) Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. *J. Virol.*, **74**, 2671–2678.
44. Welsh, P.L. and King, M.C. (2001) BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum. Mol. Genet.*, **10**, 705–713.
45. Iarovaia, O.V., Shkumatov, P. and Razin, S.V. (2004) Breakpoint cluster regions of the AML-1 and ETO genes contain MAR elements and are preferentially associated with the nuclear matrix in proliferating HEL cells. *J. Cell Sci.*, **117**, 4583–4590.
46. Kulkarni, A., Pavithra, L., Rampalli, S., Mogare, D., Babu, K., Shiekh, G., Ghosh, S. and Chattopadhyay, S. (2004) HIV-1 integration sites are flanked by potential MARs that alone can act as promoters. *Biochem. Biophys. Res. Commun.*, **322**, 672–677.
47. Johnson, C.N. and Levy, L.S. (2005) Matrix attachment regions as targets for retroviral integration. *Virol. J.*, **2**, 68.
48. Blasquez, V.C., Sperry, A.O., Cockerill, P.N. and Garrard, W.T. (1989) Protein: DNA interactions at chromosomal loop attachment sites. *Genome*, **31**, 503–509.
49. Bae, Y.S., Kawasaki, I., Ikeda, H. and Liu, L.F. (1988) Illegitimate recombination mediated by calf thymus DNA topoisomerase II in vitro. *Proc. Natl Acad. Sci. USA*, **85**, 2076–2080.
50. Keeney, S., Giroux, C.N. and Kleckner, N. (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, **88**, 375–384.
51. Girod, P.A., Zahn-Zabal, M. and Mermod, N. (2005) Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines. *Biotechnol. Bioeng.*, **91**, 1–11.
52. Urlaub, G., Kas, E., Carothers, A.M. and Chasin, L.A. (1983) Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell*, **33**, 405–412.
53. Blunt, T., Finnie, N.J., Taccioli, G.E., Smith, G.C., Demengeot, J., Gottlieb, T.M., Mizuta, R., Varghese, A.J., Alt, F.W., Jeggo, P.A. et al. (1995) Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell*, **80**, 813–823.
54. Hinz, J.M., Tebbs, R.S., Wilson, P.F., Nham, P.B., Salazar, E.P., Nagasawa, H., Urbin, S.S., Bedford, J.S. and Thompson, L.H. (2006) Repression of mutagenesis by Rad51D-mediated homologous recombination. *Nucleic Acids Res.*, **34**, 1358–1368.
55. Derouazi, M., Martinet, D., Besuchet, Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., Hacker, D.L., Beckmann, J.S. and Wurm, F.M. (2006) Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem. Biophys. Res. Commun.*, **340**, 1069–1077.
56. Milligan, L., Antoine, E., Bisbal, C., Weber, M., Brunel, C., Forne, T. and Cathala, G. (2000) H19 gene expression is up-regulated exclusively by stabilization of the RNA during muscle cell differentiation. *Oncogene*, **19**, 5810–5816.
57. Karlen, Y., McNair, A., Perseguers, S., Mazza, C. and Mermod, N. (2007) Statistical significance of quantitative PCR. *BMC Bioinformatics*, **8**, 131.
58. Akita, H., Ito, R., Kamiya, H., Kogure, K. and Harashima, H. (2007) Cell cycle dependent transcription, a determinant factor of heterogeneity in cationic lipid-mediated transgene expression. *J. Gene Med.*, **9**, 197–207.
59. Kim, J.M., Kim, J.S., Park, D.H., Kang, H.S., Yoon, J., Baek, K. and Yoon, Y. (2004) Improved recombinant gene expression in CHO cells using matrix attachment regions. *J. Biotechnol.*, **107**, 95–105.
60. Folger, K.R., Thomas, K. and Capecchi, M.R. (1985) Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Mol. Cell. Biol.*, **5**, 59–69.
61. Liang, F., Han, M., Romanienko, P.J. and Jasin, M. (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA*, **95**, 5172–5177.
62. Wyman, C., Ristic, D. and Kanaar, K. (2004) Homologous recombination-mediated double-strand break repair. *DNA Repair*, **3**, 827–833.
63. Wong, E.A. and Capecchi, M.R. (1986) Analysis of homologous recombination in cultured mammalian cells in transient expression and stable transformation assays. *Somat. Cell Mol. Genet.*, **12**, 63–72.
64. Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. and Takeda, S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.*, **17**, 5497–5508.
65. Delacote, F., Han, M., Stamato, T.D., Jasin, M. and Lopez, B.S. (2002) An *xrcc4* defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells. *Nucleic Acids Res.*, **30**, 3454–3463.
66. Chikeman, J.A., Khalsa, G.J., Barton, S.L. and Nickoloff, J.A. (2001) Homologous recombinational repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and -independent mechanisms. *Genetics*, **157**, 579–589.
67. Allen, C., Kurimasa, A., Brennehan, M.A., Chen, D.J. and Nickoloff, J.A. (2002) DNA-dependent protein kinase suppresses double-strand break-induced and spontaneous homologous recombination. *Proc. Natl Acad. Sci. USA*, **99**, 3758–3763.
68. Pierce, A.J., Hu, P., Han, M., Ellis, N. and Jasin, M. (2001) Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. *Genes Dev.*, **15**, 3237–3242.
69. Jeggo, P.A. (1997) DNA-PK: at the cross-roads of biochemistry and genetics. *Mutat Res.*, **384**, 1–14.
70. Kaufman, R.J. (2000) Overview of vector design for mammalian gene expression. *Mol. Biotechnol.*, **16**, 151–160.
71. Chusainow, J., Yang, Y.S., Yeo, J.H., Toh, P.C., Asvadi, P., Wong, N.S. and Yap, M.G. (2008) A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnology and Bioengineering*, **102**, 1182–1196.
72. Henikoff, S. (1998) Conspiracy of silence among repeated transgenes. *Bioessays*, **20**, 532–535.
73. West, S.C. (2003) Molecular views of recombination proteins and their control. *Nat. Rev. Mol. Cell Biol.*, **4**, 435–445.
74. Robins, D.M., Ripley, S., Henderson, A.S. and Axel, R. (1981) Transforming DNA integrates into the host chromosome. *Cell*, **23**, 29–39.
75. Folger, K.R., Wong, E.A., Wahl, G. and Capecchi, M.R. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol. Cell Biol.*, **2**, 1372–1387.
76. Thomas, K.R., Folger, K.R. and Capecchi, M.R. (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell*, **44**, 419–428.
77. Chen, H.H., Ho, Y.P., Jiang, X., Mao, H.Q., Wang, T.H. and Leong, K.W. (2008) Quantitative comparison of intracellular unpacking kinetics of polyplexes by a model constructed from quantum dot-FRET. *Mol. Ther.*, **16**, 324–332.
78. Jacquier, A. and Dujon, B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, **41**, 383–394.

79. Kostriken,R., Strathern,J.N., Klar,A.J., Hicks,J.B. and Heffron,F. (1983) A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell*, **35**, 167–174.
80. Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.W. (1983) The double-strand-break repair model for recombination. *Cell*, **33**, 25–35.
81. Smih,F., Rouet,P., Romanienko,P.J. and Jasin,M. (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.*, **23**, 5012–5019.
82. Chevalier,B.S. and Stoddard,B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, **29**, 3757–3774.
83. Bode,J., Kohwi,Y., Dickinson,L., Joh,T., Klehr,D., Mielke,C. and Kohwi-Shigematsu,T. (1992) Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science*, **255**, 195–197.
84. Bode,J., Winkelmann,S., Gotze,S., Spiker,S., Tsutsui,K., Bi,C., Prashanth,A.K. and Benham,C. (2006) Correlations between scaffold/matrix attachment region (S/MAR) binding activity and DNA duplex destabilization energy. *J. Mol. Biol.*, **358**, 597–613.