Review

Plant Biotechnology, 19 (4), 219–228 (2002)

Mechanism of Transgene Integration into a Host Genome by Particle Bombardment

Hiromichi MORIKAWA[1,2]*, Atsushi SAKAMOTO[1], Hiroo HOKAZONO[1], Kohei IRIFUNE[3] and Misa TAKAHASHI[1]

[1]*Department of Mathematical and Life Science, Graduate School of Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan*

[2]*Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST)*

[3]*Hiroshima Prefectural University, School of Bioresources, Shobara, Hiroshima 727-0023, Japan*

**Corresponding author* E-mail address: hmorikaw@sci.hiroshima-u.ac.jp

Received 20 August 2002; accepted 9 September 2002

Abstract

DNA sequences of transgenic loci in the transgenes of tobacco and *Arabidopsis* were analyzed. Almost all transgenic loci studied contained cis elements characteristic of nuclear matrix attachment regions (MARs). A MAR sequence isolated from one such locus of tobacco cells was cloned into the original transformation vector, and introduced into tobacco cells. The presence of the MAR sequence resulted in a five- to tenfold increase in the transformant yields, and it appeared to stimulate both the expression of the transgene and its integration into the host genome. A search using the BLAST program revealed that the transgenic loci contained junction region sequences that may have originated from different chromosomes of the nuclear genome and from the chloroplast genome.

Key words: Arabidopsis, integration, junction region, nuclear matrix attachment region (MAR), particle bombardment, recombination, nuclear scaffold attachment region (SAR), tobacco, transformation, transgene locus

Abbreviations

MAR, nuclear matrix attachment region; SAR, nuclear matrix attachment region.

Introduction

Sequence rearrangement and recombination of duplicated genes has been postulated to be an important evolutionary mechanism of the genome (Ohno, 1973; Clegg *et al.*, 1997). Recent comprehensive analyses by Blanc *et al.* (2000) and Vision *et al.* (2000) suggest that a large proportion (>60%) of the *Arabidopsis* genome results from duplication. The analysis of molecular biological and cell biological characteristics of transgenic loci will provide invaluable information not only on the mechanism of the integration of transgenes into the host genome, but also for a better understanding of the molecular evolution of the genome. This information is also indispensable for improving genetic transformation technology for recalcitrant plant species.

We have been studying methods of direct gene transfer to plants (Hashimoto *et al.*, 1985; Morikawa *et al.*, 1986, 1989; Iida *et al.*, 1990a,b), including the characterization of transgenics obtained by such methods (Iida *et al.*, 1991; Yamashita *et al.*, 1991; Nishihara *et al.*, 1993; Sawasaki *et al.*, 1994; Takahashi and Morikawa, 1996; Takahashi *et al.*, 2001) and the mechanism of transgene integration into the host genome (Sawasaki *et al.*, 1998; Shimizu *et al.*, 2001b). As part of these investigations we sequenced a transgenic locus of transformed cells of tobacco, and discovered that the locus includes a characteristic genomic sequence containing various elements specific to the nuclear matrix attachment regions (MARs) (Shimizu *et al.*, 1994). We then cloned a 507-bp portion of this MAR sequence into the original transgene construct, and used particle bombardment to deliver the new construct into tobacco cells to investigate the effect of the MAR sequence on the transformation frequency. We found that the transformation frequency was increased five- to tenfold by the presence of the MAR sequence (Shimizu *et al.*,

2001b). Interestingly, the junction regions in the trangene loci of transgenic *Arabidopsis* plants also appeared to be a MAR (Sawasaki *et al.*, 1998).

In this review, we first briefly summarize the current understanding of MARs in the eukaryotic genome, and secondly describe major characteristics of transgenic loci and the junction regions in transgenic *Arabidopsis* plants and cultured tobacco cells. Thirdly, we summarize how the transformation frequency is affected by the presence of a MAR sequence from a trangene locus, and discuss possible mechanisms in relation to the results in plants and animals reported by previous authors.

Structure of eukaryotic genome DNA

The genomic DNA is considered to be highly organized in the eukaryotic nucleus, and much of this organization is attributed to a nonhistone, proteinaceous scaffold or matrix to which the 30-nm chromatin fiber is anchored through nuclear scaffold attachment regions (SARs) or MARs at every 50-200 kb to form loop-shaped functional domains (Cockerill and Garrard, 1986; Lewin, 1997). MARs are DNA elements that are 300 bp to several thousand base-pairs long, which are operationally defined by their affinity for the nuclear matrix. MARs exhibit various characteristic motifs, such as adenine-thymine (AT) tracts (Koo *et al.*, 1986), base-unwinding elements (Benham *et al.*, 1997; Bode *et al.*, 2000), and recognition sites for the type-II topoisomerase (Topo II) from *Drosophila* (Sander and Hsieh, 1985). MARs occur exclusively in eukaryotic genomes (Bode *et al.*, 1995, 1996, 2000).

It has been calculated that DNA elements that bind to the nuclear matrices of HeLa cells comprise approximately 2% of the entire genome (Jackson *et al.*, 1996). Assuming that the ratio of MAR sequences in the genome is the same as that of HeLa cells and that their length is 1 kb, the number of S/MARs in the *Arabidopsis* genome (10^8 bp/haploid genome) is approximately 2000. The length of the loop domain is therefore thought to be approximately 49 kb. In maize, the locations of matrix attachment sites along a contiguous 280-kb region on chromosome 1 exhibit nine potential loops that vary in length from 6 to >75 kb (Avramova *et al.*, 1995).

It has been proposed that MARs are involved in several genome functions, such as stimulation and normalization of the expression integrated reporter genes (Allen *et al.*, 1993; Thompson *et al.*, 1994, 1995; Wang *et al.*, 1996; Jenuwein *et al.*, 1997; Sandhu *et al.*, 1998), and insulating transgenes from "position effects" (Eissenberg and Elgin, 1991; Breyne *et al.*, 1992; Wang *et al.*, 1996). MARs may also function as the origins of replication (Gasser, 1991), and be involved in chromosome condensation (Strick and Laemmli, 1995). At present, however, the exact function of MARs in vivo is not well understood. Computational analyses (Benham *et al.*, 1997; Bode *et al.*, 2000) have indicated that MARs have base-unpairing elements which destabilize the DNA duplex in response to stress. Topo II, for which MARs have a high affinity, introduces double-stranded breaks of DNA (Sperry *et al.*, 1989). Therefore, it is logical to propose that MARs are targets for the integration of transgenes into the host genome, given that plant and mammalian MARs function similarly.

Junction regions in the transgenic loci from Arabidopsis

Three transgenic plant lines (Sawasaki *et al.*, 1994; Takahashi and Morikawa, 1996) of *Arabidopsis* (designated A, B, and C) that bear chimeric genes of *bar* (phosphinothricin acetyltransferase) or *hph* (hygromycin phosphotransferase) were analyzed. All of these lines had a single Southern hybridization (SH) band of the *bar* or *hph* gene (Sawasaki *et al.*, 1998). The total DNA was isolated from leaves of aseptically grown T2 plants, and junction regions were amplified by the inverse polymerase chain reaction (IPCR) (Ochman *et al.*, 1988) as reported previously (Sawasaki *et al.*, 1998). The nucleotide sequences in the transgenic loci of lines A, B, and C (designated as loci A, B, and C, respectively) were determined, and deposited in the DDBJ database (accession nos. AB003139, AB003140, AB003141, AB003142). Based on these results, the structure of the transgenic locus in each line was reconstructed as shown in **Fig. 1** (Sawasaki *et al.*, 1998). Two, three, and four junction regions were identified in loci A, B, and C, respectively: 789 and 1259 bp, designated A1 and A2; 1074, 907, and 1111 bp, designated B1, B2, and B3; and 480, 106, 482, and 527 bp, designated C1, C2, C3, and C4; respectively.

All the junction regions except C2 and C4 had a high AT content (65-76%). In addition, all the junction regions except C2 and C4 were rich in MAR motifs, including the A and T boxes (Gasser and Laemmli, 1986) and *Drosophila* Topo-II cleavage consensus sites (Sander and Hsieh, 1985). AT-rich sequences carrying MAR motifs therefore appear to be characteristic of junction regions. The junction regions appeared to have a propensity for curvature (Sykes *et al.*, 1988), which is a characteristic of MARs (Sawasaki *et al.*, 1998). These findings suggest that the integration of transgenes

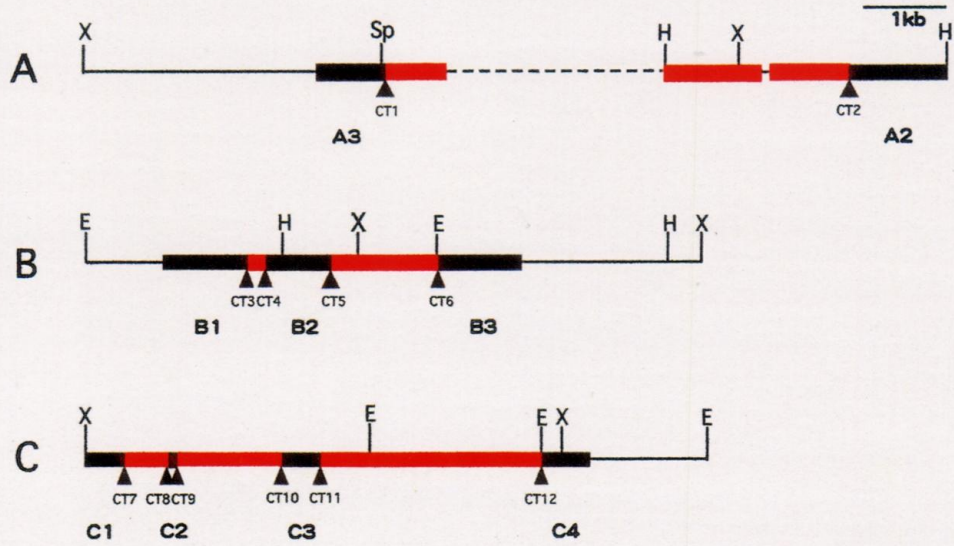

Fig. 1 Structure of transgenic loci in transformant lines A, B, and C of *Arabidopsis thaliana*. Thick red lines indicate transgenes, black lines indicate junction regions. Broken lines indicate the region not cloned, thin lines indicate the cloned sequence that was not determined. Arrowheads (CT1 to CT12) show chromosome/transgene (CT) junction positions. X, *Xba*I ; Sp, *Spe*I ; H, *Hin*dIII ; E, *Eco*RI. (Redrawn and modified after Sawasaki *et al*., Gene, Elsevier, Amsterdam, 1998.)

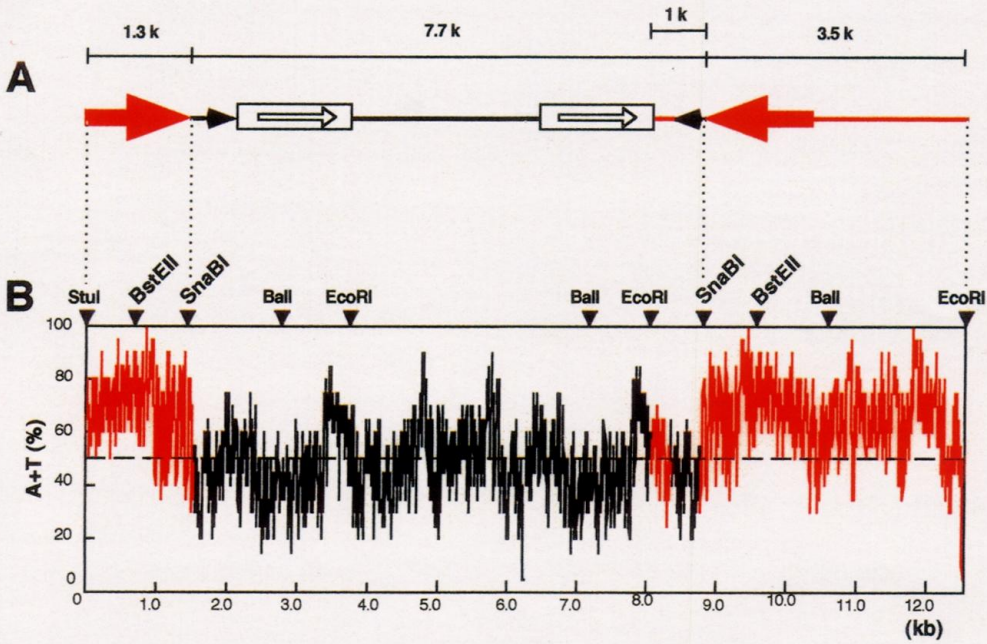

Fig. 2 Schematic model of a sequenced 12.5-kb fragment of the transgene/junction region (A), and its adenine-thymine (AT) content (B), as cloned from a transgenic line (line 2-1) of tobacco BY-2 cells. (A) The 7.7-kb core fragment consists mostly of transgenes flanked by a 1.3-and 3.5-kb genomic sequences at the 5' and 3' ends, respectively. The 7.7-kb core has a complete copy of pCaMVNEO (4.4kb), its partial sequence, and a 1-kb sequence of the genomic DNA and pUC vector sequence. The black line and black arrows in the upper panel indicate the pUC vector sequence in the transgene. Boxed arrows correspond to the expression cassette for the *npt*II gene. The red line and red arrows show the genomic sequence. Note that the identical 1.3-kb genomic sequences and 0.65-kb pUC vector sequences are oriented inversely on the 5' and 3' sides of the core. Restriction sites are indicated by arrowheads. (B) AT contents of the genomic DNA and transgenes are shown by red and black tracings, respectively. (After Shimizu *et al*., The Plant Journal, Blackwell, Oxford, 2001.).

by particle bombardment tends to occur at AT-rich regions carrying MAR motifs rather than at random regions in the genome. An in vitro binding assay of the junction regions was performed - using the method of Hall *et al.* (1991) - against nuclear matrices isolated from tobacco. DNA fragments carrying a 350- to 700-bp portion of each junction region were cut with the appropriate restriction enzyme(s) from the respective IPCR-amplified fragments. Binding efficiency was quantified as a percentage: $P_{RA}/(S_{RA}+P_{RA})$ x 100, where PRA and SRA are the radioactivities of the fragments in the pellet and supernatant fractions, respectively. Junction region fragments had a binding efficiency of 32 ± 7 to 87 ± 8% (mean ± SD), higher than the value for yeast autonomously replicating sequence which represents a positive control (24 ± 2%). These quantitative results confirm our conclusion that all these junction regions carry a MAR (Sawasaki *et al.*, 1998).

BLAST search of the junction regions

The junction region sequences were compared with those from GenBank, EMBL, DDBJ, and PDB databases on the NCBI web site (http://www.ncbi.nlm.nih.gov/), by using the BLASTN program (Schaffer *et al.*, 2001) to find possible identities. The results are summarized in **Table 1**. The entire

Table 1 Identity of junction regions to nuclear and chloroplast genomes of *Arabidopsis thaliana*

Junction region (accession no.)	Length (bp)	Position	Identity (%)	Position in nuclear genome clone or chloroplast genome clone of *Arabidopsis* (accession no.)
Junction regions that showed identity with nuclear genome				
A1 (AB003139)	789	1-778	96.8% (757/782)	93996-93224 in BAC clone F5A8 of chromosome 1 (AC004146)
A2 (AB003140)	1259	998-1075	100% (78/78)	106134-106211 of YAC contig no. 73 of chromosome 4 (AL161577)
		1101-1191	100% (91/91)	106237-106327 of YAC contig no. 73 of chromosome 4 (AL161577)
		1203-2256	99.7% (1054/1057)	106329-107385 of YAC contig no.73 of chromosome 4 (AL161577)
B1(AB003141)	1074	28-1063	98.1% (1022/1042)	155455-156495 of YAC contig no. 58 of chromosome 4 (AL161558)
B2 (AB003141)	907	1212-1448	95.0% (226/238)	70859-70622 of P1 clone MIO24 of chromosome 5 (AB010074)
		1543-2121	98.7% (573/580)	76796-76218 of P1 clone MSG15 of chromosome 5 (AB015478)
B3 (AB003141)	1111	3365-3909	99.2% (541/545)	64043-64585 of YAC contig no. 59 of chromosome 4 (AL161559)
		3927-4057	100% (131/131)	64604-64734 of YAC contig no. 59 of chromosome 4 (AL161559)
		4116-4475	98.3% (354/360)	64801-65159 of YAC contig no. 59 of chromosome 4 (AL161559)
C4 (AB003142)	527	78-527	99.3% (447/450)	24122-23673 of BAC clone T2K12 of chromosome 5 (AL590346)
Junction regions that showed identity with chloroplast genome				
C1 (AB003142)	480	1-310	100% (310/310)	977-668 of ndhG gene of chloroplast genome (X99278)
		307-330	100% (24/24)	642-665 of ndhG gene of chloroplast genome (X99278)
C2 (AB003142)	106	1020-1041	100% (22/22)	689-668 of ndhG gene of chloroplast genome (X99278)
		1038-1057	100% (20/20)	642-661 of ndhG gene of chloroplast genome (X99278)
C3 (AB003142)	482	2333-2496	100% (164/164)	66879-66716 of chloroplast genome (NC_000932)
		2508-2730	100% (223/223)	66699-66477 of chloroplast genome (NC_000932)

sequence of junction region A1 showed a 96.8% identity with a BAC genomic clone of chromosome 1, and a similar result was obtained when region A1 was compared with a BAC clone F1N21 (accession no. AC002130) of chromosome 1 (data not shown). On the other hand, a 1223-bp sequence of region A2 showed almost 100% identities with three different fragments of chromosome 4 (**Table 1**), and a similar result was obtained when region A2 was compared with a BAC clone T10C21 (accession no. AL109787) of chromosome 4 (data not shown). Thus, junction region A1 in locus A may have originated from chromosome 1, while region A2 in the same locus may have originated from chromosome 4. Since line A showed a single Southern band for the (*bar*) transgene (Sawasaki *et al.*, 1998), this locus must be located on a single chromosome. The chromosome on which locus A is located requires further investigation on details of nucleotide sequences of the locus.

Region B1 showed a 98.1% identity with a fragment of chromosome 4. Similar results were obtained when region B1 was compared with BAC clones F21P8 (accession no. AL022347) and F7H19 (accession no. AL031018) of chromosome 4 (data not shown). In contrast, region B2 showed high identities with two P1 clones that are known to be present in tandem on chromosome 5. Region B3 showed 98-100% identities with three different fragments of chromosome 4 (**Table 1**), and similar results were obtained when region B3 was compared with a BAC clone F16G20 (accession no. AL031326) of chromosome 4 (data not shown). Thus, two junction regions of the transgenic locus B may have originated from chromosome 4, while one of the junction regions in the same locus may have originated from chromosome 5. The entire length of the locus including these three regions, as cloned by IPCR, was no more than 5 kb (Sawasaki *et al.*, 1998). It is probable that the corresponding fragments of chromosomes 4 and 5 are spatially located close to each other in the nucleus, and that the chromosome fragments had been first connected with the transgene by recombination and then integrated thereafter into the host genome. Whether the locus B is located on chromoseom 4 or 5 is an intriguing and very important subject for future study.

Region C4 showed a 99.3% identity with a fragment of chromosome 5. To our surprise, regions C1, C2, and C3 did not show any identity with the nuclear genome, but showed identity with the chloroplast genome of *Arabidopsis* (**Table 1**). Region C1 showed 100% identities with two different fragments of the ndhG gene of chloroplast, region C2 showed 100% identities with two different fragments of the same gene, and region C3 showed 100% identities with two different fragments of the chloroplast genome (**Table 1**). Bècause hygromycin-resistant plants of line C segregated at a ratio of 37 to 9 (resistant to sensitive plants), the transgene is thought to be integrated not in the chloroplast genome but in the nuclear genome (Sawasaki *et al.*, 1998). Kikuchi *et al.* (1987) showed changes in the copy number of nuclear DNAs between the embryo and callus, and that the nuclear DNAs of the tissues contain chloroplast DNA. Thorsness and Fox (1990) showed that mitochondrial DNA in yeast escapes from the mitochondria to the nucleus. Moreover, 10% of yeast transformants are reported to have mitochondrial DNA insertions at their junctions (Schiestl *et al.*, 1993). More recently, the Arabidopsis project group showed that plastid DNA fragments are present in the nuclear genome, and suggested that the transfer of DNAs is still occurring between these two genomes (The Arabidopsis Genome Initiative, 2000). Probably, DNA fragments derived from the chloroplast *ndhG* gene that have escaped from the chloroplast to the nucleus are integrated within the junction region (see **Table 1**). Identification of the chromosome bearing this locus having chloroplast genome fragments by using fluorescence *in situ* hybridization (FISH) analysis is to be investigated.

Junction regions in a transgenic locus in the genome of cultured tobacco cells

Fig. 2 shows the schematic model of a transgene/junction region cloned from a transgenic tobacco cell line. The total DNA from a transformant line (line 2-1) of tobacco (Iida *et al.*, 1990a) obtained by particle bombardment with pCaMV-NEO (Fromm *et al.*, 1986) was digested with *Stu*I (which has no restriction sites in the introduced transgene). pCaMVNEO contains the neomycin phosphotransferase II gene (*npt* II) under the control of the cauliflower mosaic virus (CaMV) 35S promoter and the nopaline synthase polyadenylation signal, which is flanked to pUCPiAN7 which comprises the polylinker PiAN7 inserted into pUC8 (Fromm *et al.*, 1986). The digested DNA was fractionated by sucrose-density-gradient centrifugation according to the method of Sambrook *et al.* (1989). A 12.5-kb DNA fragment bearing junction regions was cloned into λ FIX II, and the entire 12.5-kb sequence determined (DDBJ accession no. D84238). This transgenic locus had a 7.7-kb core sequence consisting mostly of transgenes and genomic sequences flanking the 5' and 3' ends of the core (**Fig. 2A**). The 7.7-kb core had a complete

copy of plasmid pCaMVNEO at the 5' end, flanked by a partial sequence with a complete expression cassette for the *npt* II gene and a 1-kb sequence. Integration of the transgene concatemer to the genomic DNA has been shown (Riggs and Bates, 1986). The 1-kb sequence carried a 650-bp pUC-PiAN7 (a portion of the transgene mentioned above) fragment which was identical (but inversely oriented) to the fragment at the 5' end of the core. The remaining 350-bp sequence did not seem to have any homology to the pCaMVNEO sequence. We therefore concluded that it was derived from the genomic DNA of tobacco, which suggests that rearrangement had occurred between the integrated pUCPiAN7 and genomic DNA in the transformant. A similar result - that integrated transgenic DNA is interspersed by host DNA - has been reported by Pawlowski and Somers (1998).

Genomic DNA fragments flanking the 5' and 3' sides of the core were, respectively, about 1.3- and 3.5-kb long. These genomic sequences contained more AT than did the core sequences (see **Fig. 2B**). The 1.3-kb portion directly connected to the 3' end of the core of the 3.5-kb fragment had an inverse orientation (see **Fig. 2A**) that was identical to the 1.3-kb fragment of the 5' side, except for five bases at positions 332 (deletion), 414, 452, 469, and 841 (base replacement) (Shimizu *et al.*, 2001b).

The 1.3-kb sequence has several A boxes, T boxes (Gasser and Laemmli, 1986), and unwinding elements (Bode *et al.*, 1992), together with a Topo-II cleavage site (Sander and Hsieh, 1985), all of which are characteristics of a MAR (Bode *et al.*, 1995, 2000). A 507-bp *Sna*BI/*Bst*EII fragment excised from the 1.3-kb sequence on the 5' side was designated TJ1, cloned into pUC19, and studied further.

An in vitro binding assay of the TJ1 sequence was performed against the nuclear scaffold isolated from tobacco cells as described above. Approximately 65% of the total radioactivity of TJ1 was recovered in the pellet fraction, with the remaining appearing in the supernatant fraction, which provided confirmation that TJ1 is a MAR (Shimizu *et al.*, 2001b). One recognizable feature of bent DNA is its slower mobility at a low temperature during polyacrylamide gel electrophoresis (Koo *et al.*, 1986; Diekmann, 1987). The electrophoretic mobility of TJ1 changes with the temperature of the electrophoresis gel (Shimizu *et al.*, 2001b), a clear indication that TJ1 has bending characteristics that are characteristic of MARs (Homberger, 1989), which again confirms that TJ1 is a MAR.

The insertion of transgenes into the host genome is a multistep process consisting of different types of recombination (Bode *et al.*, 1995). The integration of transgenes into eukaryotes has been postulated to occur by a random, nonhomologous recombination between the exogenous DNA and genomic DNA (Lewin, 1997), and thereby no consensus sequences in transgenic loci are thought to be present in the genome. However, the results of the analyses on the transgenic loci of *Arabidopsis* and tobacco as presented here indicate that transgenes are located not randomly in the host genome but rather are flanked with AT-rich MAR sequences. The presence of AT-rich sequences near junction regions has also been reported with transformants obtained by the calcium phosphate method (Takano *et al.*, 1997) and *Agrobacterium*-mediated transformation (Gheysen *et al.*, 1987, 1990). The presence of a MAR close to an insertion site of transferred DNA (T-DNA) containing a chloramphenicol acetyltransferase gene in a transgenic petunia obtained using the *Agrobacterium* method (Dietz *et al.*, 1994). In this transformation, the single-stranded T-DNA cut from tumor-inducing plasmid, with which a virulent protein VirD2 (encoded by the tumor-inducing plasmid in bacteria) is covalently attached at the 5' end, is delivered into the plant nucleus, and the integration of the single-stranded T-DNA is initiated by association between plant DNA and the 3' end of the single-stranded T-DNA. This step is followed by the ligation of the 5' end of the single-stranded T-DNA to the 3' end of the plant DNA (Tinland, 1996). The precise nature of the plant DNA is not yet clear. MARs have also been found in the integration sites for retroviral vectors in mammalian cells (Mielke *et al.*, 1996). Retroviruses are a family of single-stranded RNA viruses that replicate through a DNA intermediate. The viral DNA is initially found in the cytoplasm of the host cell, assembled as a 160 S integrase complex which contains all enzymatic functions for the integration reaction except for the activities required for the terminal stages that are provided by the repair system of the host cell, residing at the nuclear matrix. The fact that the transgenic loci in the transformants obtained by these different systems in plants and animals contains a MAR may not be a simple coincidence, but indicative that direct gene transfer, and the *Agrobacterium* and retroviral methods share common mechanisms in the integration of transgenes, irrespective of whether a double strand or single strand is being integrated into the eukaryotic genome.

The nuclear matrix is reported to play an important role in the regulation of nuclear metabolic functions, and matrix attachment is thought to be

critical in genome replication, and in integration and recombination in mammalian cells (Brylawski *et al.*, 2000; Shimizu *et al.*, 2001a). One of the major structural components of the nuclear matrix in mammalian cells is Topo II, which is reported to relax supercoiled DNA by introducing double-stranded breaks, passages, and resealing (Sperry *et al.*, 1989; Bode *et al.*, 1995). It is conceivable that Topo II plays a key role in the integration of transgenes into the host genome. However, among 12 chromosome/transgene junctions in transgene loci of *Arabidopsis* (see **Fig. 1**), 10 had the cleavage sites for type-I topoisomerase I rather than Topo II near the junction point (Sawasaki *et al.*, 1998), thus Topo II might not be the only topoisomerase that is involved in the integration of transgenes.

Sperry *et al.* (1989) were the first to show that MARs are found at recombination sites of mammalian cells, and proposed that MARs are "hotspots" for the illegitimate recombination in eukaryotes. MARs are DNA elements that are operationally defined by their affinity for the nuclear matrix, and it is possible that certain MARs make open chromatin, and therefore make the host genome more accessible to transgenes.

The fact that a transgenic locus contains genomic fragments originating from different chromosomes and the organellar genome (**Table 1**) suggests that recombination between transgene DNA and nuclear (and organellar) DNA occurs first at somewhere on the nuclear matrix, and then whole or partial recombinants are integrated into the host genome.

The insertion of transgenes into MAR areas may be preferentially mediated by the Topo-II/topoisomerase-I mechanism described above. Alternatively, the integration itself may occur at random, but transformants with the respective transgene being integrated into MAR areas may be recovered as a consequence of MAR-mediated stimulation of gene expression of the marker gene during the *in vitro* selection process. In oncogenic human viruses, the viral integration in tumor cells - which is associated with progression of malignancy - is reported to occur near MARs of the host genome, while integrations into "nontumor" cells show no significant correlation with these regions (Shera *et al.*, 2001). Similarly, the integration of a mammalian virus (the woodchuck hepatitis virus) is reported to occur in the MAR locus of the woodchuck genome (D'Ugo *et al.*, 1998; Bruni *et al.*, 1999). It is conceivable that the MAR-mediated enhancement of the expression of tumor virus genes or MAR-mediated stimulation of integration, or both, resulted in the increased yield of tumor cells in those experiments. This hypothesis could be tested by examining how MARs isolated directly from transgenic loci affect the integration frequency.

Effect of a MAR on transformation frequency

To study the effect of the TJ1 MAR on transformant yields, TJ1 sequences were inserted into pCaMVNEO at both the 5' and 3' ends of the expression cassette of the *npt* II gene (the resulting plasmid was designated pTJ1NEO). pCaMVNEO and pTJ1NEO were introduced separately into tobacco BY-2 cells by microprojectile bombardment. Transformants were selected as described elsewhere (Iida *et al.*, 1990a). The transformation frequency was increased five- to tenfold by the insertion of the TJ1 MAR into the *npt* II expression cassette as reported previously (Shimizu *et al.*, 2001b). Southern blot analysis showed that all the Geneticin-resistant calli contained the npt II gene (data not shown). It is conceivable that MAR sequences enhance integration of the transgene into genomic DNA, or that MAR sequences enhance transgene expression which thereby increases transformation efficiency, or both (Shimizu *et al.*, 2001b).

To narrow down these possibilities, the NPTII enzymatic activity per copy of the *npt* II gene in the pTJ1NEO transformants was quantified and compared with that in pCaMVNEO transformants. The copy number of the *npt* II transgene in transformant lines bearing pCaMVNEO or pTJ1NEO was determined by the quantitative PCR method, which used the total DNA of transformants (four lines each of the pCaMVNEO and pTJ1NEO transformants) as templates, and primers that define the 1239-bp fragment between the CaMV 35S promoter and *npt* II coding region and those that define the 746-bp fragment in the *npt* II coding region. The 1239-bp product was stained with ethidium bromide (EtBr), and the intensity of the band was measured by densitometry (EtBr method). The 746-bp product was hybridized with a ^{32}P-labeled fragment of the *npt* II gene, and the intensity of the band measured with an image analyzer (SH method). In general, the EtBr method gave a higher gene copy number than the SH method. Both methods, however, clearly showed that the copy number of the transgenes per tobacco genome was 3-5 times lower in the pTJ1NEO transformants than in the pCaMVNEO transformants. In addition, variations in the copy number were somewhat less for the former than for the latter (4 ± 1 and 14 ± 6 by the EtBr method and 2 ± 1 and 4 ± 1 by the SH method, for the pTJ1NEO and pCaMVNEO transformants, respectively). This suggests that the bombardment of plasmid vectors, into which MAR sequences are inserted, produces transformants with low and uniform copy numbers.

NPT II enzyme activities in those eight transformant lines and in nontransformed control cells were measured, from which the activity of the enzyme per gene copy was estimated. The pTJ1NEO transformants appeared to exhibit an *npt* II activity per gene copy that was approximately 5 times higher than that for the pCaMVNEO transformants (Shimizu *et al.*, 2001b). The relative variation was somewhat smaller in the former than the latter transformants, at 270 ± 140 and 54 ± 39, respectively. The conclusion is that TJ1 has the typical functions of a MAR: normalization and stimulation of the expression of integrated transgenes. We are currently studying the effect of TJ1 using the gene for green fluorescent protein (Chiu *et al.*, 1996) as a nonselective reporter in order to avoid the possibility of selection pressure interfering with the transformation frequency. This study may enable us to distinguish the effect of TJ1 on transient expression from that on recombination/integration event.

The present study clearly shows that TJ1 is very useful in the production of transgenic plants, which is consistent with previous results showing that MAR-based transformation vectors improve transformation frequency in plants (Meyer *et al.*, 1988; Buising and Benbow, 1994; Galliano *et al.*, 1995) and animals (Menck *et al.*, 1998; Gutiérrez-Adán and Pintado, 2000), and that MARs stimulate the expression of integrated transgenes (Allen *et al.*, 1993; Thompson *et al.*, 1994, 1995; Wang *et al.*, 1996; Jenuwein *et al.*, 1997; Sandhu *et al.*, 1998).

Meyer *et al.* (1988) have shown that a 2-kb genomic DNA fragment (designated TBS) cloned from *Petunia hybrida* increases the transformation frequency. A 517-bp fragment taken from TBS was reported to contain a DNA-unwinding element and bind to the nuclear scaffold (Buising and Benbow, 1994; Galliano *et al.*, 1995). These characteristics of the TBS fragment from *Petunia hybrida* are similar to our TJ1 fragment isolated from a transgenic locus of tobacco. However, no other clearcut similarities in DNA sequences were present between petunia TBS and the 1.3-kb junction region sequence (containing TJ1) cloned from tobacco.Ohno (1973) proposed that whole-genome duplication is an important evolutionary mechanism, and Clegg *et al.* (1997) proposed that recombination is a pervasive force at all levels of plant evolution. Recent comprehensive analyses by Blanc *et al.* (2000) and Vision *et al.* (2000) suggest that a large proportion (>60%) of the *Arabidopsis* genome results from duplication. These studies suggest endogenous recombination-enhancing DNA elements are important to shaping plant genomes. It is conceivable that MARs such as TJ1, which is located in a transgenic locus and increases transformation frequency, contribute to DNA rearrangement and the evolution of plant genomes.

Investigation of junction sequences of the pTJ1NEO transformants will be an intriguing subject of future study. Whether or not TJ1 sequences in the transformation vector have triggered homologous recombination in tobacco also will be an important subject. Furthermore, elucidation of heterologous effects of TJ1 sequence on the transformation and gene expressoin in those organisms that have been frequently reported to cause strong gene silencing, and in those that are recalcitrant to transformation also is an important future subject. Comprehensive analyses of junction regions in such transformants will be vital to elucidate the transgene integration mechanism in eukaryotes and to develop novel high performance transformation vectors.

Acknowledgments

This work was supported in part by Grants-in-Aid for Scientific Research (nos. 11460156 and 12025224) from the Ministry of Education, Science, Culture and Sports, Japan. Financial support from the program Research for the Future (no. JSPS-RFTF97R16001) from the Japan Society for the Promotion of Science is greatly appreciated.

References

Allen, G. C., Hall, G. E., Jr, Childs, L. C., Weissinger, A. K., Spiker, S., Thompson, W. F., 1993. Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. Plant Cell, **5**: 603-613.

Avramova, Z., SanMiguel, P., Georgieva, E., Bennetzen, J. L., 1995. Matrix attachment regions and transcribed sequences within a long chromosomal continuum containing maize Adh1. Plant Cell, **7**: 1667-1680.

Benham, C., Kohwi-Shigematsu, T., Bode, J., 1997. Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. J. Mol. Biol., **274**: 181-196.

Blanc, G., Barakat, A., Guyot, R., Cooke, R., Delseny, M., 2000. Extensive duplication and reshuffling in the Arabidopsis genome. Plant Cell, **12**: 1093-1101.

Bode, J., Benham, C., Knopp, A., 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.

Bode, J., Kohwi, Y., Dickinson, L., Joh, Y., Klehr, D., Mielke, C., Kohwi-Shigematsu, T., 1992. Biological significance of unwinding capability of nuclear matrix-associating DNAs. Science, **255**: 195-197.

Bode, J., Schlake, T., Ríos-Ramírez, M., Mielke, C., Stengert, M., Kay, V., Klehr-Wirth, D., 1995. Scaffold/matrix-attached regions: structural properties creating transcripitionally active loci. Int. Rev. Cytol., **162A**: 389-454.

Bode, J., Stengert-Iber, M., Kay, V., Schlake, T., Dietz-

Pfeilstetter, A., 1996. Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. Crit. Rev. Eukaryot. Gene Expr., **6**: 115-138.

Breyne, P., van Montagu, M. V., Depicker, A., Gheysen, G., 1992. Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. Plant Cell, **4**: 463-471.

Bruni, R., D'Ugo, E., Giuseppetti, R., Argentini, C., Rapicetta, M., 1999. Activation of the N-myc2 oncogene by woodchuck hepatitis virus integration in the linked downstream b3n locus in woodchuck hepatocellular carcinoma. Virology, **257**: 483-490.

Brylawski, B. P., Cohen, S. M., Cordeiro-Stone, M., Schell, M. J., Kaufman, D. G., 2000. On the relationship of matrix association and DNA replication. Crit. Rev. Eukaryot. Gene Expr., **10**: 91-99.

Buising, C. M., Benbow, R. M., 1994. Molecular analysis of transgenic plants generated by microprojectile bombardment: effect of petunia transformation booster sequence. Mol. Gen. Genet., **243**: 71-81.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., Sheen, J., 1996. Engineered GFP as a vital reporter in plants. Curr. Biol., **6**: 325-330.

Clegg, M. T., Cummings, M. P., Durbin, M. L., 1997. The evolution of plant nuclear genes. Proc. Natl. Acad. Sci. U. S. A., **94**: 7791-7798.

Cockerill, P. N., Garrard, W. T., 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell, **44**: 273-282.

Diekmann, S., 1987. Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. Nucleic Acids Res., **15**: 247-265.

Dietz, A., Kay, V., Schlake, T., Landsmann, J., Bode, J., 1994. A plant scaffold attached region detected close to a T-DNA integration site is active in mammalian cells. Nucleic Acids Res., **22**: 2744-2751.

D'Ugo, E., Bruni, R., Argentini, C., Giuseppetti, R., Rapicetta, M., 1998. Identification of scaffold/matrix attachment region in recurrent site of woodchuck hepatitis virus integration. DNA Cell Biol., **17**: 519-527.

Eissenberg, J. C., Elgin, S. C. R., 1991. Boundary functions in the control of gene expression. Trends Genet., **7**: 335-340.

Fromm, M.E., Taylor, L.P., Walbot, V., 1986. Stable transformation of maize after gene transfer by electroporation. Nature, **319**: 791-793.

Galliano, H., Muller, A. E., Lucht, J. M., Meyer, P., 1995. The transformation booster sequence from *Petunia hybrida* is a retrotransposon derivative that binds to the nuclear scaffold. Mol. Gen. Genet., **247**: 614-622.

Gasser, S. M., 1991. Replication origins, factors and attachment sites. Curr. Opin. Cell Biol., **3**: 407-413.

Gasser, S. M., Laemmli, U. K., 1986. Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster. Cell, **46**: 521-530.

Gheysen, G., Herman, L., Breyne, P., Gielen, J., Van Montagu, M., Depicker, A., 1990. Cloning and sequence analysis of truncated T-DNA inserts from *Nicotiana tabacum*. Gene, **94**: 155-163.

Gheysen, G., Montagu, M. V., Zambryski, P., 1987. Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA sequences. Proc. Natl. Acad. Sci. U. S. A., **84**: 6169-6173.

Gutiérrez-Adán, A., Pintado, B., 2000. Effect of flanking matrix attachment regions on the expression of microinjected transgenes during preimplantation development of mouse embryos. Transgenic Res., **9**: 81-89.

Hall, G., Jr., Allen, G. C., Loer, D. S., Thompson, W. F., Spiker, S., 1991. Nuclear scaffolds and scaffold-attachment regions in higher plants. Proc. Natl. Acad. Sci. U. S. A., **88**: 9320-9324.

Hashimoto, H., Morikawa, H., Yamada, Y., Kimura, A., 1985. A novel method for transformation of intact yeast cells by electroinjection of plasmid DNA. Appl. Microbiol. Biotechnol., **21**: 336-339.

Homberger, H. P., 1989. Bent DNA is a structural feature of scaffold-attached regions in *Drosophila melanogaster* interphase nuclei. Chromosoma, **98**: 99-104.

Iida, A., Morikawa, H., Yamada, Y., 1990a. Stable transformation of cultured tobacco cells by DNA-coated gold particles accelerated by gas-pressure-driven particle gun. Appl. Microbiol. Biotechnol., **33**: 560-563.

Iida, A., Seki, M., Kamada, M., Yamada, Y., Morikawa, H., 1990b. Gene delivery into cultured plant cells by DNA-coated gold particle accelerated by a pneumatic particle gun. Theor. Appl. Genet., **80**: 813-816.

Iida, A., Yamashita, T., Yamada, Y., Morikawa, H., 1991. Efficiency of particle-bombardment-mediated transformation is influenced by cell cycle stage in synchronized cultured cells of tobacco. Plant Physiol., **97**: 1585-1587.

Jackson, D.A., Bartlett, J., Cook, P.R., 1996. Sequences attaching loops of nuclear and mitochondrial DNA to underlying structures in human cells: the role of transcription units. Nucleic Acids Res., **24**: 1212-1219.

Jenuwein, T., Forrester, W. C., Fernández-Herrero, L. A., Laible, G., Dull, M., Grosschedl, R., 1997. Extension of chromatin accessibility by nuclear matrix attachment regions. Nature, **385**: 269-272.

Kikuchi, S., Takaiwa, F., Oono, K., 1987. Variable copy number DNA sequences in rice. Mol. Gen. Genet., **210**: 373-380.

Koo, H. S., Wu, H. M., Crothers, D. M., 1986. DNA bending at adenine thymine tracts. Nature, **320**: 501-506.

Lewin, B., 1997. Genes VI. Oxford University Press, New York.

Menck, M., Mercier, Y., Campion, E., Lobo, R. B., Heyman, Y., Renard, J. P., Thompson, E. M., 1998. Prediction of transgene integration by noninvasive bioluminescent screening of microinjected bovine embryos. Transgenic Res., **7**: 331-341.

Meyer, P., Kartzke, S., Niedenhof, I., Heidmann, I., Bussmann, K., Saedler, H., 1988. A genomic DNA segment from *Petunia hybrida* leads to increased transformation frequencies and simple integration patterns. Proc. Natl.

Acad. Sci. U. S. A., **85**: 8568-8572.

Mielke, C., Maass, K., Tummler, M., Bode, J., 1996. Anatomy of highly expressing chromosomal sites targeted by retroviral vector. Biochemistry, **35**: 2239-2252.

Morikawa, H., Iida, A., Matsui, C., Ikegami, M., Yamada, Y., 1986. Gene transfer into intact plant cells by electroinjection through cell walls and membranes. Gene, **41**: 121-124.

Morikawa, H., Iida, A., Yamada, Y., 1989. Transient expression of foreign genes in plant cells and tissues obtained by a simple biolistic device (particle-gun). Appl. Microbiol. Biotechnol., **31**: 320-322.

Nishihara, M., Ito, M., Tanaka, I., Kyo, M., Ono, K., Irifune, K., Morikawa, H., 1993. Expression of the β-glucuronidase gene in pollen of lily (Lilium longiflorum), tobacco (Nicotiana tabacum), Nicotiana rustica, and peony (Paeonia lactiflora) by particle bombardment. Plant Physiol., **102**: 357-361.

Ochman, H., Gerber, A. S., Hartl, D. L., 1988. Genetic applications of an inverse polymerase chain reaction. Genetics, **120**: 621-623.

Ohno, S. 1973. Ancient linkage group and frozen accidents. Nature, **244**: 259-262.

Pawlowski, W. P., Somers, D. A., 1998. Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. Proc. Natl. Acad. Sci. U. S. A., **95**: 12106-12110.

Riggs, C. D., Bates, G. W., 1986. Stable transformation of tobacco by electroporation: evidence for plasmid concatenation. Proc. Natl. Acad. Sci. U. S. A., **83**: 5602-5606.

Sambrook, J., Fritsch, E. F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd Edn-Cold Spring Harbor Laboratory Press, Plainview, New York.

Sander, M., Hsieh, T.-S., 1985. Drosophila topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. Nucleic Acids Res., **13**: 1057-1072.

Sandhu, J. S., Webster, C. I., Gray, J. C., 1998. A/T-rich sequences act as quantitative enhancers of gene expression in transgenic tobacco and potato plants. Plant Mol. Biol. **37**: 885-896.

Sawasaki, T., Seki, M., Anzai, H., Irifune, K., Morikawa, H., 1994. Stable transformation of Arabidopsis with the bar gene using particle bombardment. Transgenic Res., **3**: 279-286.

Sawasaki, T., Takahashi, M., Goshima, N., Morikawa, H., 1998. Structures of transgene loci in transgenic *Arabidopsis* plants obtained by particle bombardment: junction regions can bind to nuclear matrices. Gene, **218**: 27-35.

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V. and Altschul, S. F., 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res., **29**: 2994-3005.

Schiestl, R. H., Dominska, M., Peters, T. D., 1993. Transformation of Saccharomyces cerevisiae with nonhomologous DNA: illegitimate integration of transforming DNA into yeast choromosomes and in vivo ligation of transforming DNA to mitochondrial DNA sequences. Mol. Gen. Genet., **13**: 2697-2705.

Shera, K. A., Shera, C. A., McDougall, J. K., 2001. Small tumor virus genomes are integrated near nuclear matrix attachment regions in transformed cells. J. Virol., **75**: 12339-12346.

Shimizu, K., Kawakami, S., Yamada, T., Goshima, N., Tanaka, T., Irifune, K., Morikawa, H., 1994. Determination of DNA sequences in a transgenic locus in transgenic tobacco cells obtained by particle bombardment. In the abstract of the Annual Meeting of the Japanese Society of Plant Physiologists, 1994, Tsukuba, p. 138 (in Japanese).

Shimizu, N., Miura, Y., Sakamoto, Y., Tsutsui, K., 2001a. Plasmids with a mammalian replication origin and a matrix attachment region initiate the event similar to gene amplification. Cancer Res., **61**: 6987-6990.

Shimizu, K., Takahashi, M., Goshima, N., Kawakami, S., Irifune, K., Morikawa, H., 2001b. Presence of an SAR-like sequence in junction regions between an introduced transgene and genomic DNA of cultured tobacco cells: its effect on transformation frequency. Plant J., **26**: 375-384.

Sperry, A. O., Blasquez, V. C., Garrard, W. T., 1989. Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. Proc. Natl. Acad. Sci. U. S. A., **86**: 5497-5501.

Strick, R., Laemmli, U. K., 1995. SARs are *cis* DNA elements of chromosome dynamics: synthesis of a SAR repressor protein. Cell, **83**: 1137-1148.

Sykes, R. C., Lin, D., Hwang, S. J., Framson, P. E., Chinault, A. C., 1988. Yeast ARS function and nuclear matrix association coincide in a short sequence from the human HPRT locus. Mol. Gen. Genet., **212**: 301-309.

Takahashi, M., Morikawa, H., 1996. High frequency stable transformation of Arabidopsis by particle bombardment. J. Plant Res., **109**: 331-304.

Takahashi, M., Sasaki, Y., Ida, S., Morikawa, H., 2001. Enrichment of nitrite reductase gene improves the ability of Arabidopsis thaliana plants to assimilate nitrogen dioxide. Plant Physiol., **126**: 731-741.

Takano, M., Egawa, H., Ikeda, J.-E., Wakasa, K., 1997. The structures of integration sites in transgenic rice. Plant J., **11**: 353-361.

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, **408**: 796-815.

Thompson, E. M., Adenot, P., Tsuji, F. I., Renard, J.-P., 1995. Real time imaging of transcriptional activity in live mouse preimplantation embryos using a secreted luciferase. Proc. Natl. Acad. Sci. U. S. A., **92**: 1317-1321.

Thompson, E. M., Christians, E., Stinnakre M.-G., Renard, J.-P., 1994. Scaffold attachment regions stimulate HSP70.1 expression in mouse preimplantation embryos

but not in differentiated tissues. Mol. Cell. Biol., **14**: 4694-4703.

Thorsness, P. E., Fox, T. D., 1990. Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. Nature, **284**: 426-430.

Tinland, B., 1996. The integration of T-DNA into plant genome. Trends Plant Sci., **1**: 178-184.

Vision, T. J., Brown, D. G., Tanksley, S. D., 2000. The origins of genomic duplications in Arabidopsis. Science, **290**: 2114-2-17.

Wang, D.-M., Taylor, S., Levy-Wilson, B., 1996. Evaluation of the function of the human apolipoprotein B gene nuclear matrix association regions in transgenic mice. J. Lipid Res., **37**: 2117-2124.

Yamashita, T., Iida, A., Morikawa, H., 1991. Evidence that more than 90% of β-glucuronidase-expressing cells after particle bombardment directly receive the foreign gene in their nucleus. Plant Physiol., **97**: 829-831.