The Plant Journal (1994) 6(6), 957–968

TECHNICAL ADVANCE

Reduced variation in transgene expression from a binary vector with selectable markers at the right and left T-DNA borders

Madan K. Bhattacharyya*, Bruce A. Stermer and Richard A. Dixon
The Samuel Roberts Noble Foundation, Plant Biology Division, P.O. Box 2180, Ardmore, Oklahoma 73402, USA

Summary

A new binary vector for *Agrobacterium*-mediated plant transformation was constructed, in which two selectable markers, for kanamycin and hygromycin resistance, were placed next to the right and left T-DNA borders, respectively, and a CaMV 35S promoter-driven β-glucuronidase (GUS) gene was placed between these markers as a reporter gene (transgene). Using double antibiotic selection, all transgenic tobacco plants carrying at least one intact copy of the T-DNA expressed the transgene, and this population exhibited reduced variability in transgene expression as compared with that obtained from the parent vector pBI121. Absence of the intact transgene was the major reason for transgenic plants with little or no transgene expression. Integration of truncated T-DNAs was also observed among transgenic plants that expressed the transgene and carried multiple T-DNA inserts. The copy number of fully integrated T-DNAs was positively associated with transgene expression levels in R_0 plants and R_1 progeny populations. Variability due to position effect was determined among 17 plants carrying a single T-DNA insert. The coefficient of variability among these plants was only 35.5%, indicating a minor role for position effects in causing transgene variability. The new binary vector reported here can therefore be used to obtain transgenic populations with reduced variability in transgene expression.

Introduction

Agrobacterium tumefaciens-mediated gene transfer is the most commonly used method of plant transformation. The desired gene is placed within the T-DNA of the bacterium and transferred to the plant through infection of the plant tissues (Gheysen *et al.*, 1989). The T-DNA integrates into the plant genome, and integration is stable and inherited in a Mendelian fashion (Weising *et al.*, 1988). This method is used widely in the molecular analysis of plant genes, particularly for functional promoter analysis such as identification of *cis*-elements. The expression level of a transgene may, however, vary up to one 100-fold or more (An, 1986; Jones *et al.*, 1985; Peach and Velten, 1991), and thus the interpretation of quantitative data becomes very difficult. This variation in transgene expression level may be the outcome of more than one factor.

The level of expression of a transgene has been considered to be influenced by the position in the genome into which the gene is integrated. The influence of the genome on transgene expression is known as the 'position effect' (Jaenisch *et al.*, 1981). In addition to this position effect-dependent variation, the expression level of a transgene can be influenced further by the number of integrated copies of the gene (Gendloff *et al.*, 1990; Hobbs *et al.*, 1990, 1993; Jones *et al.*, 1987; Leisy *et al.*, 1989; Schuch *et al.*, 1989).

André *et al.* (1986) presented evidence for the integration of T-DNA into transcribed DNA sequences. In more than 30% of the cases studied, the integration of T-DNAs takes place into transcriptionally active regions of the genome (Herman *et al.*, 1990; Koncz *et al.*, 1989). Integration of T-DNAs into transcriptionally active chromatin can result in insertion into endogenous promoters located in such regions. Based on this principle, attempts have been made to develop promoter-trap vectors for the cloning of genes based on the tissue- or organ-specific expression the promoters impart on a reporter gene (André *et al.*, 1986). T-DNA tagging mutagenesis is similarly based on integration of T-DNA into a gene or its promoter (Feldmann, 1991). Conversely, fusion of T-DNA on to an active endogenous promoter may interfere with the expression of the genes present in the T-DNA. Depending upon the orientation of the integrated gene to that of the endogenous promoter to which the gene is fused, promoter occlusion or transcriptional interference phenomena may inhibit transgene expression (Proudfoot, 1986) and thus contribute to variability in transgene expression (Breyne *et al.*, 1992a; Ingelbrecht *et al.*, 1991).

Received 29 April 1994; revised 12 August 1994; accepted 5 September 1994.
*For correspondence (fax +1 405 221 7380).

A significant component of transgene variability may also result from the integration of partial T-DNA carrying a non-intact transgene (Dean *et al.*, 1988a; Jones *et al.*, 1987). We have constructed a new binary vector, pNFHK1, to eliminate this part of the variability in transgene expression. In this vector, the transgene is placed between two selectable antibiotic resistance markers located next to the right and left borders. The promoters of the two selectable markers are next to the T-DNA borders. Selection for both markers should yield transformants exhibiting integration of intact T-DNA and potentially reduced transcriptional interference, since any active transcription from endogenous plant promoters should presumably be terminated at the polyadenylation signals of either marker gene. Analysis of R_0 and R_1 transgenic tobacco populations using this vector indicates reduced variability of transgene expression and reduced position effects. This vector may therefore be advantageous for functional promoter analysis using stable transformation strategies, as well as for other plant transformation studies.

Results

Construction of a binary vector and variability of transgene expression among R_0 transgenic tobacco plants

The binary vector, pNFHKI, was constructed as shown in Figure 1.

Tobacco leaf discs were co-cultivated (Horsch *et al.*, 1985) with *Agrobacterium tumefaciens* LBA4404 carrying pBI121 or pNFHK1. The infected leaf discs were transferred to MS medium containing either kanamycin sulfate (only kanamycin sulfate for pBI121), hygromycin or both kanamycin sulfate and hygromycin, and transformed calli were selected. Apical bud tissues of individual regenerated transgenic tobacco plants were analyzed for GUS activity. Among 22 transgenic plants carrying pBI121, 18 plants showed significant GUS activity (Figure 2a) the rest showing only 1% or less of the mean GUS activity of all 22 plants. Among the 31 transgenic plants carrying pNFHK1 obtained after selection for resistance to kana-

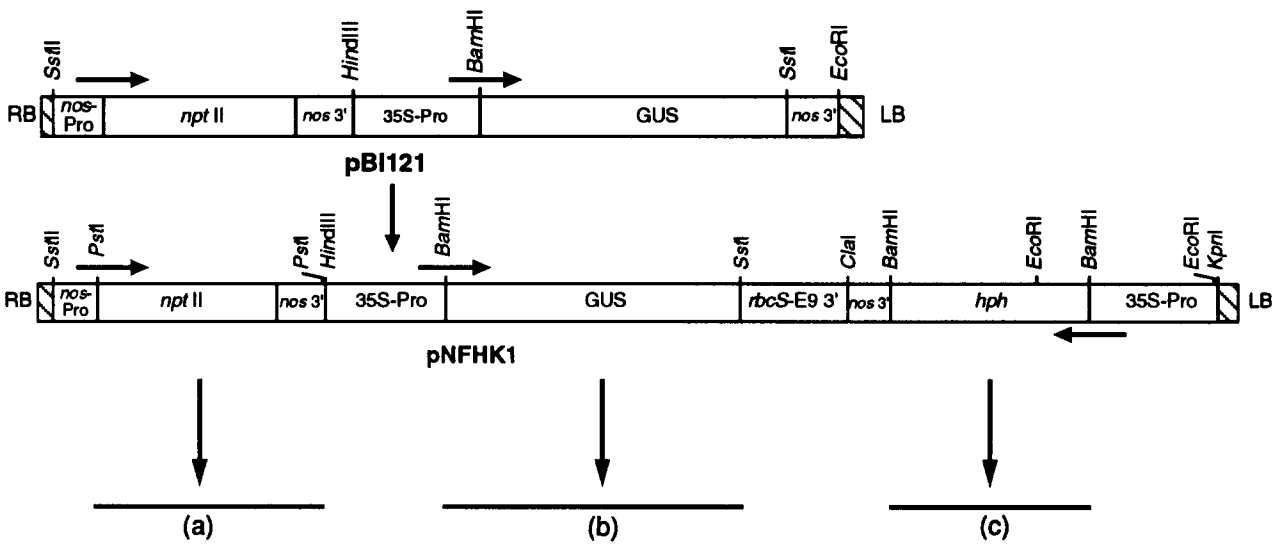


Figure 1. Construction of a binary vector with different selectable markers at the left and right T-DNA borders and probes that were used for Southern blot analysis.

The binary vector plasmid, pNFHK1, was constructed as follows. The binary vector pBI121 (Jefferson *et al.*, 1987) was cut to completion with *Eco*RI, end-filled with the large fragment of *Escherichia coli* DNA polymerase I (Klenow fragment), and *Kpn*I linkers added by blunt-end ligation. The hygromycin-resistance gene, *hph*, was excised from the binary vector pOCA19 by digestion with *Kpn*I and *Bsp*106; the approximately 3 kb *Kpn*I–*Bsp*106 fragment of pOCA19 contains the cauliflower mosaic virus (CaMV) 35S promoter, the *hph* coding region and the nopaline synthase (*nos*) polyadenylation signal. The above fragment was cloned into pBlueScript (SK^-) to yield the plasmid pNF1. The fragment carrying the *rbcS*-E9 polyadenylation signal was obtained from a 0.8 kb *Xho*I–*Cla*I fragment of the binary vector pKYLX71 (Berger *et al.*, 1989), which was cloned into pBlueScript (SK^-) to obtain pNF2. The *Sst*I-*Hin*dIII fragment of pNF2 containing the *rbcS*-E9 3' polyadenylation signal was isolated and cloned into pNF1 to obtain pNF3. The *Hin*dIII site was eliminated from pNF3 by digesting with *Hin*dIII followed by end-filling, ligation and transformation, to yield plasmid pNF4. The *Kpn*I–*Sst*I fragment carrying the *hph* gene and the *rbcS*-E9 polyadenylation signal was isolated from pNF4, and cloned into the *Sst*I–*Kpn*I sites of the *Kpn* linker-modified pBI121. The resulting plasmid-carrying parts of pBI121, the *hph* gene and the *rbcS*-E9 polyadenylation signal for polyadenylation of CaMV 35S promoter-driven β-glucuronidase (GUS) transcripts is termed pNFHK1. The following fragments of pNFHK1 were used on probes in Southern blot analysis: (a) a *Pst*I fragment including the *npt-II* gene and *nos* 3' end; (b) a *Bam*HI–*Sst*I fragment carrying the GUS gene; (c) a *Bam*HI fragment carrying the *hph* gene.

mycin alone, only 21 showed significant GUS activity (Figure 2b), the rest showing less than 0.3% of the mean GUS activity of all 31 plants. The distribution pattern of these plants for the level of GUS expression was comparable with that for transformation with the binary vector pBI121 (Figure 2a and b). In contrast, 30 out of 32 plants selected for hygromycin resistance showed significant levels of GUS activity (Figure 2c), with the two remaining plants showing less than 1.2% of the mean GUS activity of all the hygromycin-selected resistant plants.

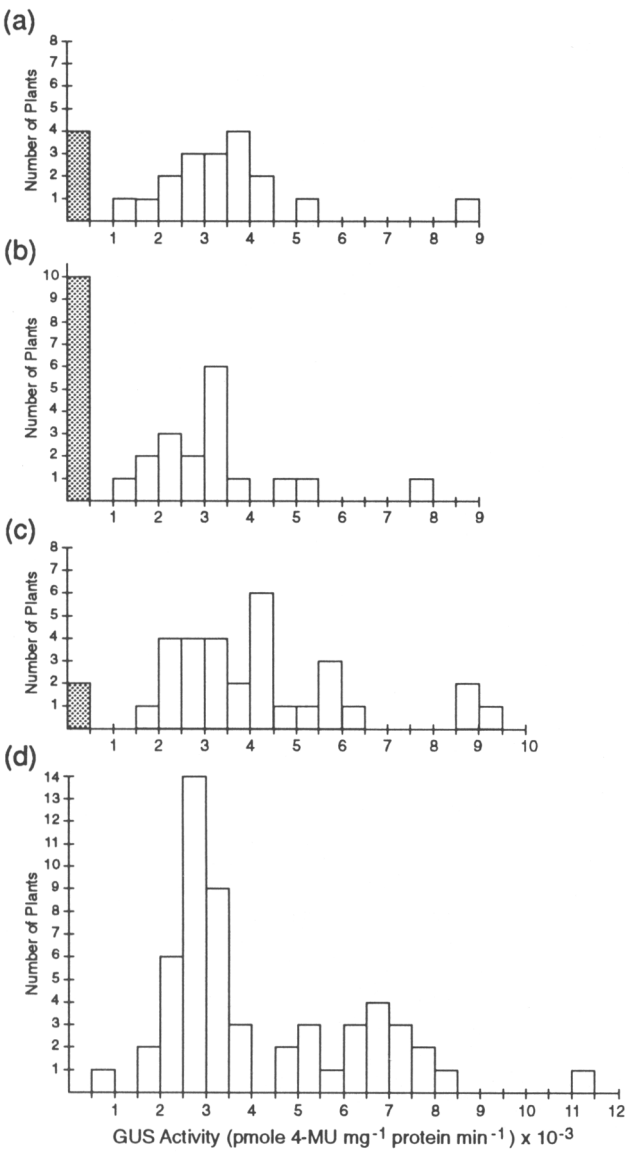


Figure 2. GUS activity in apical bud tissue of transformants obtained using the binary vectors pBI121 (a), or pNFHK1 (b, c and d). Transformants were obtained after selection in medium containing kanamycin sulfate (a and b), hygromycin (c), or both kanamycin sulfate and hygromycin (d). The dotted bars in (a), (b), and (c) represent the classes of transformants showing less than 2% mean GUS activity. These plants were termed GUS-negative.

A sample of 55 pNFHK1 transgenic tobacco plants selected for resistance to both antibiotics was analyzed for GUS expression. All of these plants showed significant levels of GUS activity (Figure 2d). The variability for levels of GUS activity was analyzed in natural logarithmic scale (Nap *et al.*, 1993) (Table 1). Significantly reduced variance values were recorded for R_0 plants obtained from transformation with pNFHK1 and selection in media containing both hygromycin and kanamycin (Figure 2d) as compared with those calculated for R_0 plants obtained from transformation with pBI121 (Figure 2a) and pNFHK1 (Figure 2b) and selection in medium containing kanamycin or hygromycin (Figure 2c, for pNFHK1 only). Transgenic plants with very little GUS activity (less than 2% of the mean of the population) were regarded as GUS-negative plants. Southern blot analysis of eight of the GUS-negative plants obtained after transformation with pNFHK1 and selection on kanamycin alone, using probes derived from pNFHK1 (Figure 1) revealed that five of these plants carried the intact *npt-II* gene (Table 2). However, in two of the GUS-negative plants, the GUS gene was only partially integrated into the genome (Table 2).

Association of T-DNA copy number with GUS activity

A random sample of 25 R_0 plants covering the entire GUS-expression range of the 55 transgenic plants resistant to both antibiotics (Figure 2d) was analyzed for T-DNA copy number. DNA from each of these 25 plants was digested with *Hin*dIII and DNA blots were probed with fragments of pNFHK1 carrying the *npt-II* gene (Figure 1, probe a), the GUS gene (Figure 1, probe b) or the *hph* gene (Figure 1, probe c).

From these three DNA blots, the copy number of T-DNA inserts was estimated (Table 3). We observed partial T-DNA integration in a few transformants that carried multiple T-DNA insertions. In these transformants some of the copies did not contain both selectable markers along with the GUS gene. For example, in plant HK105 there are three T-DNA copies each carrying the GUS gene but only two of them carrying the *npt-II* gene, because one band is not large enough to account for the *npt-II* gene. This conclusion was supported by subsequent segregation analysis, in which progeny carrying this small band were kanamycin sensitive. Similarly, plant HK213 also showed a small *npt-II* hybridizing band, suggesting incomplete integration. Comparision of band patterns for plant HK203 shows differences in the number of bands hybridizing to the GUS and *hph* gene probes. This indicates partial transfer of either GUS or *hph* in some of the T-DNAs in plant HK203. Anomalies in the number of bands hybridizing to these three genes were also observed for plants HK106, HK103, HK104, HK215,

Table 1. Variation of R_0 transgenic tobacco for GUS expression[a]

Construct	Selection[b]	Natural logarithmic scale				
		Sample size (n)	Mean ($\bar{x}$)	SE	Variance[c]	RT mean[d]
pBI121	Kanamycin	22	7.25	0.41	3.75	1408
pNFHK1	Kanamycin	31	6.21	0.49	7.72	498
	Hygromycin	32	7.98	0.22	1.49*†	2922
	Kanamycin and hygromycin	55	8.21	0.07	0.25**†#	3678

[a]GUS activity is expressed as pmol methylumbelliferone (4-MU) mg^{-1} protein min^{-1}, and the mean is calculated for all the transgenic plants of an individual treatment.
[b]The MS-agar medium was amended with kanamycin sulfate (100 μg ml^{-1}), hygromycin (25 μg ml^{-1}), or both kanamycin sulfate and hygromycin.
[c]The test of equality of two variances was carried out according to Sokal and Rohlf (1981).
†Denotes significant differences at $P = 0.005$ when compared with the variances of the R_0 plants obtained from transformation with the pNFHK1 and selection in medium containing kanamycin only.
*and ** Denote significant differences at $P = 0.025$ and $P = 0.005$, respectively, when compared with the variance of the R_0 plants obtained from transformation with pBI121.
# Denotes significant differences at $P = 0.005$ when compared with the variance of the R_0 plants obtained from transformation with pNFHK1 and selection in medium containing hygromycin only.
[d]RT mean denotes retransformed mean to the original scale of measurement (GUS activity as in [a] above).

Table 2. T-DNA integration in 'GUS-negative' plants[a]

Plant no.	GUS activity[b]	Hybridizing band[e]			
		npt-II[c]	GUS[c]	Intact *npt-II* band[d]	Copy no.[e]
K-18	11.1	+	+	+	0
K-26	6.9	+	–	–	0
K-2	44.1	+	+	+	0
K-15	22.9	+	–	+	0
K-12	1.6	+	–	–	0
K-102	2.72	+	–	+	0
K-25	7.7	+	–	+	0
K-24	50.9	+	–	–	0

[a]GUS-negative plants carrying T-DNA of pNFHK1 were obtained from selection in medium containing kanamycin sulfate only (Figure 2b).
[b]GUS activity is expressed as pmol methylumbelliferone (4-MU) mg^{-1} protein min^{-1}.
[c]Genomic DNA was digested with *Pst*I and probed with probe (a) (Figure 1) for *npt-II* or digested with *Hin*dIII and *Eco*RI and probed with probe (b) (Figure 1) containing the GUS gene: +, presence of a band; –, absence of a band.
[d]Genomic DNA was digested with *Pst*I and probed with the *npt-II* gene and *nos* 3' end probe (probe (a) in Figure 1): +, presence of a *Pst*I fragment carrying the *npt-II* gene; –, presence of partial integration due to loss of one *Pst*I site (higher bands than predicted size).
[e]The copy number for the GUS gene is designated as zero in all cases due to the absence of an intact GUS gene fragment (*Hin*dIII to *Eco*RI fragment of pNFHK1, see Figure 1) in the Southern blot analysis.

HK211, HK200, HK213, HK202, HK205, HK204 and HK115 (Table 3). Therefore, the number of bands that hybridized to both GUS and *hph* genes was considered as the copy number of transgene insertions in R_0 plants. This analysis revealed that there were 10 independent R_0 plants, obtained by double antibiotic selection, carrying a single T-DNA, and that all of these plants showed similar levels of GUS activity. We also observed that the plants with higher levels of GUS activity contained more than one T-DNA insert. There would therefore appear to be

Table 3. Analysis of GUS activity and T-DNA copy number in R_0 transformants[a]

Plant no.	GUS activity[b]	Hybridizing band			Copy no.[c]
		npt-II	GUS	*hph*	
HK105	850	3	3	3	3
HK106	2,320	2	1	2	1
HK103	2,377	2	1	1	1
HK116	2,941	1	1	1	1
HK114	2,899	1	1	1	1
HK104	2,560	5	1	1	1
HK109	2,809	1	1	1	1
HK208	2,588	1	1	1	1
HK12	3,243	1	1	1	1
HK207	3,164	1	1	1	1
HK215	3,106	2	1	1	1
HK211	3,632	3	2	2	2
HK200	4,508	2	2	3	2
HK213	5,431	6	4	3	3
HK202	5,415	4	2	2	2
HK113	5,870	2	2	2	2
HK201	6,177	2	2	2	2
HK203	6,042	7	7	6	5
HK102	6,694	4	4	4	4
HK205	6,993	3	2	2	2
HK101	7,034	2	2	2	2
HK204	7,610	3	3	3	2
HK210	7,536	3	3	3	3
HK107	8,085	3	3	3	3
HK115	11,376	5	5	4	4

[a]R_0 transformants carrying T-DNA of pNFHK1 were obtained from selection in medium containing both kanamycin sulfate and hygromycin (Figure 2d). Twenty-five transformants were selected to represent the entire range of transformants presented in Figure 2d. The genomic DNA was digested with *Hin*dIII and the same blot was hybridized to probes a (*npt-II* and *nos* 3' end), b (GUS gene), or c (*hph* gene) (see Figure 1b).
[b]GUS activity in pmol 4-MU mg^{-1} protein min^{-1}.
[c]Copy number of GUS was determined from the number of bands that hybridized to both GUS and *hph* probes.

a positive correlation between transgene copy number and the levels of GUS activity. The estimated positive correlation coefficient (0.64) between the T-DNA copy number and the level of GUS activity among the 25 R_0 plants was significant at $P = 0.01$ (Figure 3a(i)).

The selfed progeny (R_1) of three categories of primary R_0 transformants representative of the whole R_0 population were analyzed for GUS activity and T-DNA copy number. The categories of R_1 plants were: (i) progeny of HK109, HK116 and HK208 which carried a single T-DNA insert; (ii) progeny of HK105 which carried three T-DNA copies but showed the lowest level of GUS expression among the R_0 plants; (iii) progeny of HK115 which carried four T-DNA copies and showed the highest level of GUS expression. The T-DNA copy number was estimated on the basis of band intensity (data not shown). The relative amount of DNA in individual lanes was determined relative to the intensity of an 8 kb non-specific band obtained on hybridization with the GUS gene and washing filters at 65°C in 2× SSC. A total of 38 R_1 progeny from HK109, HK116 and HK208 were examined and a highly significant (at $P = 0.01$) positive correlation (0.78) between T-DNA copy number, which was either two copies (homozygous) or one copy (heterozygous), and GUS activity was obtained (Figure 3a(ii)). A larger R_1 population derived from HK105 was assayed for GUS activity. Among 96 progeny, six were kanamycin sensitive. From this population, 15 plants with the highest GUS activities and 10 plants with the lowest GUS activities were analyzed for T-DNA copy number. The correlation (0.58) between the T-DNA copy number and the level of GUS activity of the selected 25 plants (Figure 3a(iii)) was significant at $P = 0.01$.

Forty-six R_1 progeny of HK115 were analyzed. We observed that bands for T-DNA copies T1 and T5 were superimposed on each other when filters with *Hin*dIII-digested DNA were hybridized with the *npt-II* probe (Figure 3b). Therefore, some plants (as shown by an asterisk in Figure 3b) have more than two T-DNA copies in bands representing T1 and T5 T-DNAs. However, such problems did not occur when the filter was hybridized with the GUS gene (Figure 3c). The T-DNA copy T2 was truncated and lacks the *hph* gene and also probably a part of the GUS gene. The R_2 progeny analysis of R_1 plant number 36 showed that the progeny carrying T2 were in fact GUS negative. Therefore, in copy number estimations, only T1, T3, T4 and T5 bands were counted. Again, the positive correlation (0.80) observed between T-DNA copy number and levels of GUS activity of these 46 R_1 plants (Figure 3a(iv)) was significant at $P = 0.01$.

Intratransformant variability and position effects

The analysis of transgene expression variability in the present investigation is based on single estimations for individual transformants. To understand how much of this variability is contributed by environmental and/or experimental error, we carried out two different experiments. First, 15 independent R_1 progeny of HK105 representing a wide variation of GUS expression levels were selected and each plant was cut into four pieces that were asexually propagated into four independent plants in four different Magenta boxes. The variability of these four plants of a single genotype is caused by environmental factors and/or experimental error. The analysis of these asexually propagated plants showed that the standard errors for these 15 R_1 plants ranged between 3.8 and 19.8% of the respective means (Figure 4a(i)). In the second experiment, three independent R_0 plants each

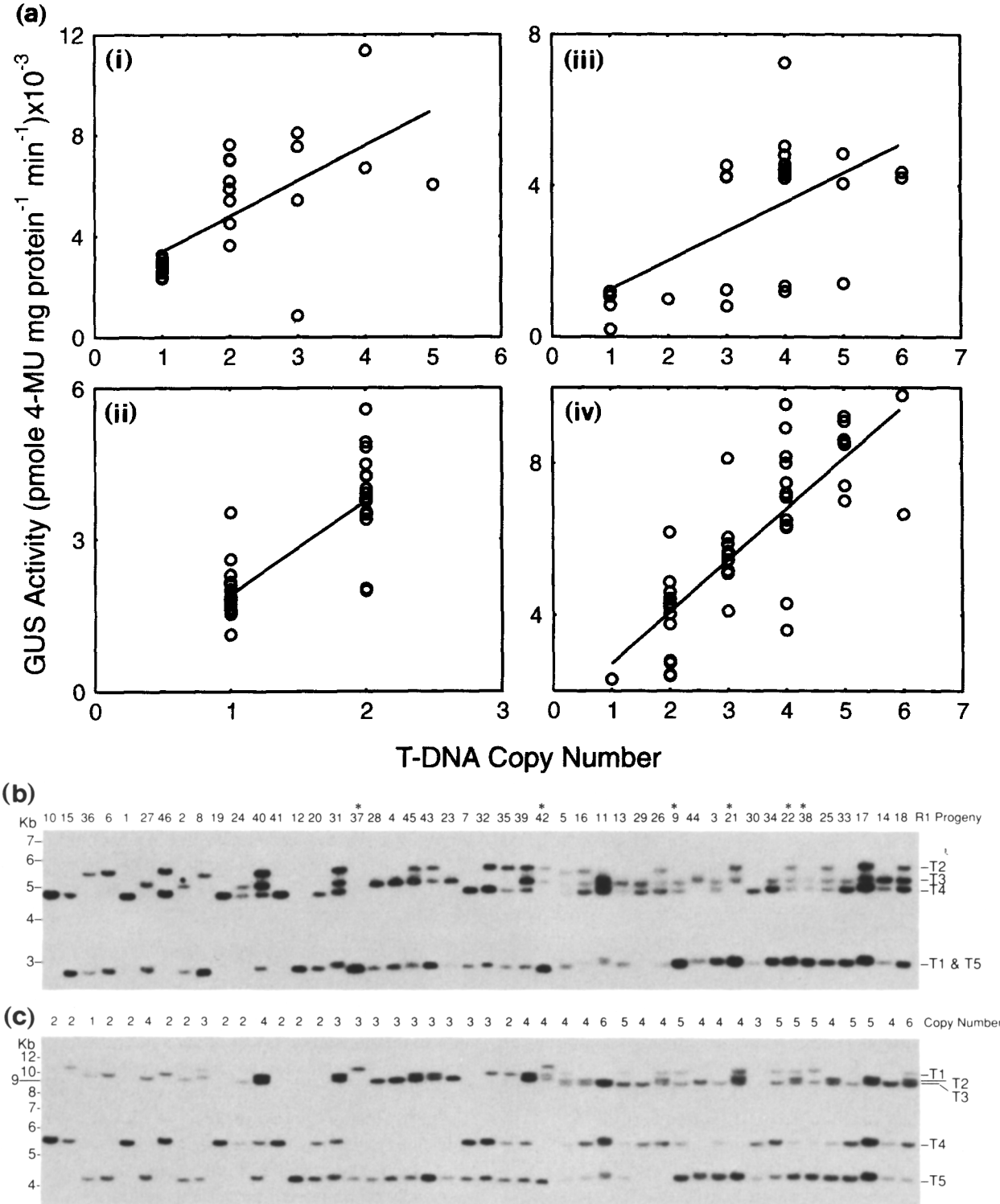
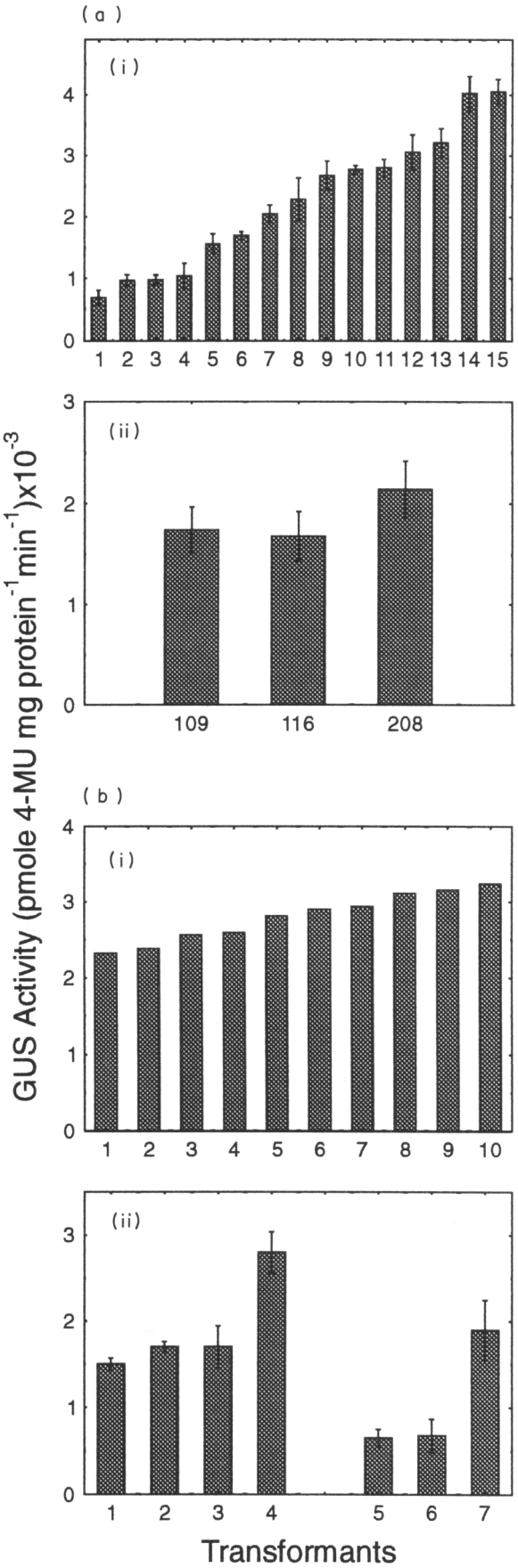


Figure 3. Relationship between T-DNA copy number and GUS activity in R_0 transformants and R_1 progeny.
(a) Shows the relationship between GUS activity and copy number among: (i) 25 R_0 transformants ($r = 0.64$); (ii) 38 R_1 progeny descended from the three R_0 plants HK109, HK116 and HK208 (see Figure 4a(ii); $r = 0.78$); (iii) 25 R_1 progeny from R_0 plant HK105 (this plant showed the lowest level of GUS expression; $r = 0.58$); and (iv) 46 R_1 progeny from the R_0 plant HK115 (which showed the highest level of GUS expression; $r = 0.8$).
(b and c) Southern blots showing copy number estimation of the individual progeny of HK115 (Figure 3a(iv)). The DNA was digested with *Hin*dIII and same filter was probed with either the *npt-II* gene (in b) or the GUS gene (in c). The progeny were arranged on the blot according to their levels of GUS activity in ascending order. The T2 T-DNA does not carry the *hph* gene. The R_2 progeny of HK115-36 carrying the T2 T-DNA showed no GUS activity. Therefore, T1, T3, T4, and T5 T-DNAs were considered for copy number determination. The R_1 progeny carrying three T-DNA copies in T1 and T5 are denoted by an asterisk in b.



carrying single, independent T-DNA inserts were taken to the R_1 generation. Mean and standard error were calculated for the individual R_1 progeny population carrying a single T-DNA copy (heterozygous) (Figure 4a(ii)). The standard errors for these three populations ranged between 12.9 and 14.4% of the respective means. In both experiments it was observed that the amount of variability in GUS expression levels caused by environmental and/or experimental error was minimal.

To determine the variability of GUS expression levels caused by the integration of T-DNA into different positions in the genome, we compared the levels of GUS activity of those transformants that contained single T-DNA inserts. Two kinds of plants were considered. First, we studied 10 primary transformants representative of the 55 primary transformants carrying a single T-DNA insert (Table 3). The GUS activity levels of these 10 plants were comparable (Figure 4b(i)). Secondly, from selfing of two primary transformants (HK105 and HK115) carrying multiple T-DNAs we obtained four classes of R_1 progeny each segregating a single T-DNA insert. For example, two progeny populations of HK115 carried T-DNAs T4 and T5. Similarly, we obtained two R_1 populations that carried the T2 and T3 T-DNAs of HK105, respectively. Three R_1 plants, two and 37 of HK115 and 18 of HK105, were selfed and from the respective R_2 populations plants carrying a single T-DNA insert (T1, T3 and T5 of HK115, and T1 of HK105) were isolated. The mean GUS activity levels along with the respective standard errors of all seven populations (four R_1 and three R_2) each carrying a single independent T-DNA insert are presented in Figure 4 (b,ii). The mean, coefficient of variation and range of GUS activity of the 10 R_0 plants and seven segregating progeny populations from two independent R_0 plants (four R_1, three R_2) with a single

Figure 4. Intratransformant and position effect-related variability in GUS expression levels.
(a) Intra-transformant variability of GUS expression levels. (i) Variability of asexually propagated clones from 15 different R_1 progeny of HK105; (ii) variability of three R_1 progeny populations carrying single T-DNA inserts that originated from primary transformants HK109(4), HK116(11) and HK208(5) that each contained a single T-DNA insert. The number in parentheses indicates the number of plants in each population.
(b) GUS activity levels of different transformants carrying single T-DNA inserts. (i) Ten independent primary transformants (R_0); (ii) Seven populations of R_1 or R_2 progeny each carrying a single T-DNA insert. (1) R_2 (three progeny) of HK115-37 carrying T1; (2) R_2 (four progeny) of HK115-2 carrying T3; (3) R_1 (10, 1, 19, and 41) of HK115 carrying T4; (4) R_1 (12) of HK115 and R_2 (two progeny) of HK115-2 carrying T5; (5) R_2 (four progeny) of HK105-18 carrying T1; (6) R_1 (86, 14, 47, 8) of HK105 carrying T2; (7) R_1 (six kanamycin-sensitive plants) of HK105 carrying T3. The number in parentheses indicates the R_1 progeny plant number. The data are mean GUS activity calculated for single-copy T-DNA (heterozygous). Values were divided by 2 if the observed plants were homozygous for the T-DNA inserts.

T-DNA were 2473 pmol 4-MU mg^{-1} protein min^{-1}, 35.5% and 648–3164 pmol 4-MU mg^{-1} protein min^{-1}, respectively.

Discussion

Transgenic variability and transcriptional interference

Wide variability in expression of a transgene among the transgenic plant population is the rule rather than the exception (An, 1986; Dean *et al.*, 1988a; Jones *et al.*, 1985; Schöffl *et al.*, 1993; Shirsat *et al.*, 1989). In most studies, a significant proportion of the transgenic plants show a low level, or absence, of expression of the transgene (see Figure 2 of Peach and Velten, 1991). As a result of this, wide ranges of expression, accounting for up to 200-fold variation, were reported in earlier studies (An, 1986; Jones *et al.*, 1985). In the present investigation, transgenic tobacco populations obtained through transformation with the binary vector pBI121, or with pNFHK1 with selection pressure for resistance to kanamycin sulfate or hygromycin only, showed a similar pattern of variability. The portion of the population of pNFHK1 transformants with no or insignificant transgene expression was eliminated when selection pressure was provided for resistance to both kanamycin sulfate and hygromycin. In this case, we observed only a 13.4-fold difference between the plants with the highest and the lowest GUS expression levels and a significant reduction in variance of ln GUS activity as compared with that for R_0 transformants obtained from either pBI121 or pNFHK1 with selection for resistance to kanamycin or hygromycin (for pNFHK1 only) (Table 1).

Southern blot analysis of the kanamycin-resistant plants with very low or no GUS activity indicated that in most of these plants the T-DNA transfer was incomplete or T-DNAs may have rearranged following transfer (Table 2). Absence of transgene DNA due to rearrangement or incomplete T-DNA transfer was found to be one cause for lack of transgene expression in previous studies (Dean *et al.*, 1988a; Jones *et al.*, 1987). Peach and Velten (1991) analyzed transgenic tobacco calli for two reporter genes, GUS and CAT, that were driven by the mannopine synthase promoter. The CAT gene was in the center, while the GUS gene was next to the left border of the T-DNA. They observed that there were more calli with no GUS activity than with no CAT activity, and that all calli with no CAT activity also showed no GUS activity. Selection for kanamycin resistance in their studies probably allowed complete integration of the CAT gene to a higher degree than the GUS gene. Other explanations, e.g. transcriptional interference from endogenous plant promoters, could not be ruled out, however. In the present study, the selection of transformants resistant to both antibiotics insured the integration of at least one intact copy of the T-DNA in a region of the genome from where both marker genes could be expressed efficiently. The placement of the GUS gene in the middle of both markers thus ensured the expression of this reporter gene in all doubly selected transformants. We observed that selection of transformants for resistance to hygromycin alone also resulted in a much higher percentage of transformants with high GUS expression levels. Selection of left borders thus appears to encourage a better T-DNA transfer than selection for right T-DNA borders. This supports the earlier conclusion that T-DNA transfer takes place from the right to the left T-DNA border (Wang *et al.*, 1984).

In addition to insuring the complete integration of the T-DNA and faithful expression of the reporter gene, the placement of the two selectable markers in the pNFHK1 vector might also have reduced potential transcriptional interference from transcriptionally active endogenous plant promoters to which the T-DNAs could be fused. It has been proposed that transcriptional interference may produce antisense RNA from a transgene, which may co-suppress the endogenous copy of the transgene (Grierson *et al.*, 1991; Mol *et al.*, 1991). Paszty and Lurquin (1990), using a transient expression system in protoplasts, showed that active transcription from a downstream promoter from the 3′ to the 5′ end of a transgene can inhibit transgene expression. It has also been shown that use of the *nos* polyadenylation signal to reduce presumed transcriptional interference significantly reduced transgenic variability (Breyne *et al.*, 1992a; Ingelbrecht *et al.*, 1991). In pNFHK1 we placed the two selectable markers facing the right and left T-DNA borders. Therefore, any transcriptional interference from endogenous promoters should be terminated following polyadenylation of RNAs transcribed from those endogenous promoters at the polyadenylation signals of the *npt-II* and *hph* genes, provided that polyadenylation and transcription termination processes in plants are coupled, as has been suggested for yeast (Osborne and Guarente, 1989; Russo and Sherman, 1989).

T-DNA copy number and transgene expression

Insertion of multiple copies of T-DNA is a common phenomenon. Expression of some or all of those copies can influence the level of expression of a transgene. Unless all the individual T-DNA inserts in a single transformant are segregated out in subsequent generations, it is difficult to assign the contribution of each of these inserts, because some intact transgenes may be silent or their expression levels may be reduced due to their integration into transcriptionally inactive heterochromatin or heterochromatin-adjacent areas, as has been shown in *Drosophila* (Dorer and Henikoff, 1994; Spofford, 1976)

and yeast (Allshire *et al.*, 1994; Gottschling *et al.*, 1990). Transgene silencing due to integration near silencer elements has also been reported (Lee and Gross, 1993).

In order to investigate the relationship between T-DNA copy number and transgene expression level we analyzed R_0 transformants and R_1 progeny from three categories of R_0 plants. The association of GUS expression levels and GUS gene copy number among the R_0 transformants was significant (Figure 3a(i)). Three independent R_0 plants containing a single T-DNA insert were selfed to obtain 38 progeny. Analysis of these R_1 progeny each carrying one (heterozygous) or two (homozygous) copies of the respective T-DNA insert revealed that the level of GUS activity positively correlated with the copy number (Figure 3a(ii)). Analysis of 25 R_1 progeny from another R_0 plant with three T-DNA copies and showing the lowest level of expression also showed that the higher expression levels were associated with the higher number of T-DNA inserts (Figure 3a(iii)). The analysis of 46 R_1 progeny from the R_0 plant carrying 4 T-DNA copies and showing the highest level of expression also showed a positive correlation between the level of GUS activity and T-DNA copy number (Figure 3a(iv)). This suggests that the high level of expression of R_0 plant HK115 was due to the cumulative effect of expression of several copies of the transgene rather than a 'position effect' causing high expression of a single copy.

The association of transgene expression level with copy number has been recently documented (Gendloff *et al.*, 1990; Hobbs *et al.*, 1993; Leisy *et al.*, 1989; Schuch *et al.*, 1989), although many other studies have failed to show any positive correlation between T-DNA copy number and transgene expression level (see review by Weising *et al.*, 1988). However, silencing of some of the T-DNA copies can result in underestimation of transgene expression per copy, making it difficult to observe the true relation between T-DNA copy number and expression level. Segregation analysis of the independent R_0 plants in the present study allowed us to overcome this problem.

Position effect and transgene expression

Among the T-DNAs that integrate into transcriptionally active areas, the transgene expression level may vary depending upon the varying combinations of enhancers and other *cis*-acting elements in the region of the genome into which the T-DNAs integrate. In transgenic mice, position-independent expression of human β-globin genes was achieved by including a large amount of upstream and downstream sequence of the globin gene (Grosveld *et al.*, 1987). Similar attempts in plants have failed to abolish apparent position-related variability (Dean *et al.*, 1988b), although, in a recent study, it was suggested that transgene variability in transgenic plants could be reduced by incorporating the chicken lysozyme matrix-associated region at the T-DNA borders (Mlynárová *et al.* 1994). Similarly, the use of plant scaffold attachment regions (SARs) has been shown to reduce transgene expression variability among stably transformed plants (Allen *et al.*, 1993; Breyne *et al.*, 1992b; Schöffl *et al.*, 1993). In the present study, we determined the levels of transgene expression for individual T-DNA inserts that were integrated into different locations in the tobacco genome. The levels of GUS expression in 10 R_0 plants carrying a single T-DNA insert were comparable (Figure 4b(i)). Analysis of segregating populations from two different R_0 plants representing the two opposite extremes of expression level and each carrying multiple T-DNA inserts allowed us to determine the levels of transgene expression for GUS activity of seven additional independent T-DNA inserts (Figure 4b(ii)). The coefficient of variability and range of GUS activity of these 10 R_0 plants and seven segregating progeny populations each carring a single independent T-DNA insert were only 35.5% and 648–3164 pmol 4-MU mg^{-1} protein min^{-1}, respectively.

On the basis of the above findings, we believe it is likely that the variability observed in a transgenic population is explained only partly by the position effect, and that T-DNA copy number may be a more important factor. Other factors, such as developmental or physiological states of tissues, can also play a significant role in transgene expression variability (Jones *et al.*, 1987). Leisy *et al.* (1989) observed that a rice glutelin promoter CAT gene fusion was expressed in tobacco seed in a gene copy number-dependent manner, whereas in non-seed tissues a wide variation in expression levels was observed. In the present investigation, we used plants that were grown under comparable growth conditions, and only analyzed the apical bud and first unfolding leaf tissues, which are the actively growing parts of the plant and can therefore be considered to be at comparable developmental and physiological states in different individuals. This may be one of the reasons why we failed to see a large variation among plants carrying individual T-DNA copies inserted at different positions of the genome (Figure 4b). In older leaves we might have encountered a larger variability. The other reason is the possibility that effects of potential enhancer or promoter activities in the regions into which the transgenes integrate were eliminated or reduced due to placement of the transgene between two selectable markers that are facing towards the outside of the T-DNA.

In summary, construction of a binary vector with selectable markers next to the right and left T-DNA borders facilitated the complete integration of at least one T-DNA molecule and expression of the transgene (CaMV 35S promoter-driven GUS gene) in each transformant.

Analysis of primary transformants and their R_1 progeny populations indicated that T-DNA copy number plays a major role in the variability of transgene expression. There was a reduced variation of transgene expression among the primary transformants and position effects appear to be a minor component of transgene expression variability. Comparison of transgene expression levels among transformants harboring a single T-DNA insert produced using different binary vectors will clarify whether the apparently minor role of position effect is a characteristic peculiar to the new binary vector reported here.

Experimental procedures

Nucleic acid manipulations

DNA manipulation for plasmid construction was as described in Sambrook *et al.* (1989). Restriction enzymes were obtained from Gibco-BRL, Stratagene and Promega. Construction of the binary vector plasmid pNFHK1 was carried out in *Escherichia coli* DH5α cells.

Plant transformation and regeneration

Agrobacterium tumefaciens LBA4404 carrying the plasmids pNFHK1 or pBI121 was used to inoculate leaf discs of *Nicotiana tabacum* cv. Xanthi (Horsch *et al.*, 1985). Transformed plants were selected on MS medium (Murashige and Skoog, 1962) containing carbenicillin (200 μg ml^{-1}) with either kanamycin sulfate (100 μg ml^{-1}), hygromycin (25 μg ml^{-1}) or both antibiotics. The transformed plants were grown axenically in Magenta boxes at a light intensity of 50 μE m^{-2} sec^{-1} with 18 h day length at 25°C.

Assay for GUS activity

Individual transformants were transferred to separate Magenta boxes containing MS medium amended with appropriate antibiotics. Transformed plants showing root formation were allowed to grow until they touched the top of the Magenta boxes, and the apical bud and first unfolding leaf were then harvested separately.

Harvested tissues from individual samples were ground in 500 μl extraction buffer (Jefferson, 1987) and soluble protein content of individual samples was determined according to Bradford (1976). Total soluble proteins varied between 250 and 500 μg per tissue sample. Determination of GUS activity in these samples was carried out by fluorimetric assay according to Jefferson (1987).

Extraction of DNA and Southern blotting

After harvesting of the apical bud and first leaf, each transformant was cut just below the fully open fourth leaf from the apex and the material above was frozen. For DNA extraction, the tissues were ground into powder, transferred into 14 ml plastic tubes, and 10 ml extraction buffer and phenol:chloroform mixture was added and mixed vigorously for at least 2 min. Following centrifugation (2500 ***g***), the nucleic acid pellets were resuspended in 400 μl of extraction buffer and transferred into Eppendorf tubes. The remainder of the steps were carried out exactly according to the protocol of White and Kaper (1989). From each sample about 50–100 μg of DNA were obtained.

DNA was cut with restriction enzymes, electrophoresed in 0.7% agarose, and transferred to nylon membranes by capillary blotting in 0.4 M NaOH, 1.5 M NaCl overnight. Membranes were neutralized in 0.5 M Tris–HCl pH 7.5 for 5 min then washed in 2× SSC and air dried. Filters were prehybridized for 1–2 h and hybridized with probe in 50% formamide buffer overnight at 42°C (Sambrook *et al.*, 1989). Washing conditions were according to the Gene Screen Plus nylon membrane protocol of DuPont. Probes were labeled according to the method of Feinberg and Vogelstein (1984).

Statistical analysis

Coefficient of variation (standard deviation over mean in percentage) and correlation coefficient were calculated according to the published method (Steel and Torrie, 1960). Significance of correlation coefficient was carried out by comparing the calculated '*r*' value with the corresponding table value of Fisher's *r* table at *n*–2 degree of freedom. Scatter diagram analysis was carried out using the GLE computer program of Information Technology Group, New Zealand.

Acknowledgments

We are grateful to Joanne Chory, Salk Institute, San Diego, for kindly providing the pOCA19 vector. We thank Cuc Ly and Ty Myers for preparing the drawings and photographs, Rick Nelson and Grant Hotter for initial review of the manuscript, and Allyson Wilkins and Connie Cabe for preparation of the manuscript.

References

Allen, G.C., Hall, Jr, G.E., Childs, L.C., Weissinger, A.K., Spiker, S. and Thomson, W.F. (1993) Scaffold atttachment regions increase reporter gene expression in stably transformed plant cells. *Plant Cell*, **5**, 601–613.

Allshire, R.C., Javerzat, J.-P., Redhead, N.J. and Cranston, G. (1994) Position effect variegation at fission yeast centromeres. *Cell*, **76**, 157–169.

An, G. (1986) Development of plant promoter expression vectors and their use for analysis of differential activity of nopaline synthase promoter in transformed tobacco cells. *Plant Physiol.* **81**, 86–91.

André, D., Colau, D., Schell, J., Van Montagu, M. and Hernalsteens, J.-P. (1986) Gene tagging in plants by a T-DNA insertion mutagen that generates APH(3')II–plant gene fusions. *Mol. Gen. Genet.* **204**, 512–518.

Berger, P.H., Hunt, A.G., Domier, L.L., Hellmann, G.M., Stram, Y., Thornbury, D.W. and Pirone, T.P. (1989) Expression in transgenic plants of a viral gene product that mediates insect transmission of potyviruses. *Proc. Natl Acad. Sci. USA*, **86**, 8402–8406.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.

Breyne, P., Gheysen, G., Jacobs, A., Van Montagu, M. and Depicker, A. (1992a) Effect of T-DNA configuration on transgene expression. *Mol. Gen. Genet.* **235**, 389–396.

Breyne, P., Van Montagu, M., Depicker, A. and Gheysen, G. (1992b) Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. *Plant Cell*, **4**, 463–471.

Dean, C., Favreau, M., Tamaki, S., Bond-Nutter, D., Dunsmuir, P. and Bedbrook, J. (1988a) Expression of tandem gene fusions in transgenic tobacco plants. *Nucl. Acids Res.* **16**, 7601–7617.

Dean, C., Jones, J., Favreau, M., Dunsmuir, P. and Bedbrook, J. (1988b) Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. *Nucl. Acids Res.* **16**, 9267–9283.

Dorer, D.R. and Henikoff, S. (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell*, **77**, 993–1002.

Feinberg, A.P. and Vogelstein, B. (1984) A technique for radiolabeling DNA endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266–267.

Feldmann, K.A. (1991) T-DNA insertion mutagenesis in *Arabidopsis:* mutational spectrum. *Plant J.* **1**, 71–82

Gendloff, E.H., Bowen, B. and Buchholz, W.G. (1990) Quantitation of chloramphenicol acetyl transferase in transgenic tobacco plants by ELISA and correlation with gene copy number. *Plant Mol. Biol.* **14**, 575–583.

Gheysen, G., Herman, L., Breyne, P., Van Montagu, M. and Depicker, A. (1989) *Agrobacterium tumefaciens* as a tool for the genetic transformation of plants. In *Genetic Transformation and Expression* (Butler L.O., Harwood, C. and Moseley, B.E.B., eds). Andover: Intercept, pp. 151–174.

Gottschling, D.E., Aparicio, O.M., Billington, B.L. and Zaklan, V.A. (1990) Position effect at *S. cerevisiae* telomeres: Reversible repression of Pol II transcription. *Cell*, **63**, 751–762.

Grierson, D., Fray, R.G., Hamilton, A.J., Smith, C.J.S. and Watson, C.F. (1991) Does co-suppression of sense genes in transgenic plants involve antisense RNA? *Trends Biotechnol.* **9**, 122–123.

Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, **51**, 975–985.

Herman, L., Jacobs, A., Van Montagu, M. and Depicker, A. (1990) Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events. *Mol. Gen. Genet.* **224**, 248–256.

Hobbs, S.L.A., Kpodar, P. and DeLong, C.M.O. (1990) The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. *Plant Mol. Biol.* **15**, 851–864.

Hobbs, S.L.A., Warkentin, T.D. and DeLong, C.M.O. (1993) Transgene copy number can be positively or negatively associated with transgene expression. *Plant Mol. Biol.* **21**, 17–26.

Horsch, R.B., Fry, J.E., Hoffmann, N.L., Eichholtz, D., Rogers, S.G. and Fraley, R.T. (1985) A simple and general method for transferring genes into plants. *Science*, 227, 1229–1231.

Ingelbrecht, I., Breyne, P., Vancompernolle, K., Jacobs, A., Van Montagu, M. and Depicker, A. (1991) Transcriptional interference in transgenic plants. *Gene*, **109**, 239–242.

Jaenisch, R., Jähner, D., Nobis, P., Simon, I., Löhler, J., Harbers, K. and Grotkopp, D. (1981) Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. *Cell*, **24**, 519–529.

Jefferson, R.A. (1987) Assaying chimeric genes in plants: The GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387–405.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) Gus gene fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Jones, J.D.G., Dunsmuir, P. and Bedbrook, J. (1985) High level expression of introduced chimaeric genes in regenerated transformed plants. *EMBO J.* **4**, 2411-2418.

Jones, J.D.G., Gilbert, D.E., Grady, K.L. and Jorgensen, R.A. (1987) T-DNA structure and gene expression in petunia plants transformed by *Agrobacterium tumefaciens* C58 derivatives. *Mol. Gen. Genet.* **207**, 478–485.

Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Körber, H., Redei, G.P. and Schell, J. (1989) High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl Acad. Sci. USA*, **86**, 8467–8471.

Lee, S. and Gross, D.S. (1993) Conditional silencing: The *HMRE* mating-type silencer exerts a rapidly reversible position effect on the yeast *HSP82* heat shock gene. *Mol. Cell Biol.* **13**, 727–738.

Leisy, D.J., Hnilo, J., Zhao, Y. and Okita, T.W. (1989) Expression of a rice glutelin promoter in transgenic tobacco. *Plant Mol. Biol.* **14**, 41–50.

Mlynárová, L., Loonen, A., Heldens, J., Jansen, R.C., Keizer, P., Stiekema, W.J. and Nap, J.-P. (1994) Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. *Plant Cell*, **6**, 417–426.

Mol, J., van Blokland, R. and Kooter, J. (1991) More about co-suppression. *Trends Biotechnol.* **9**, 182–183.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.

Nap, J.-P., Keizer, P. and Jansen, R. (1993) First-generation transgenic plants and statistics. *Plant Mol. Biol. Rep.* **11**, 156–164.

Osborne, B.I. and Guarente, L. (1989) Mutational analysis of a yeast transcriptional terminator. *Proc. Natl Acad. Sci. USA*, **86**, 4097–4101.

Paszty, C.J.R. and Lurquin, P.F. (1990) Inhibition of transgene expression in plant protoplasts by the presence in *cis* of an opposing 3'-promoter. *Plant Sci.* **72**, 69–79.

Peach, C. and Velten, J. (1991) Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. *Plant Mol. Biol.* **17**, 49–60.

Proudfoot, N.J. (1986) Transcriptional interference and termination between duplicated α-globin gene constructs suggests a novel mechanism for gene regulation. *Nature*, **322**, 562–565.

Russo, P. and Sherman, F. (1989) Transcription terminates near the poly(A) site in the *CYC1* gene of the yeast *Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA*, **86**, 8348–8352.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schöffl, F., Schröder, G., Kliem, M. and Rieping, M. (1993) An SAR sequence containing 395 bp DNA fragment mediates enhanced, gene-dosage-correlated expression of a chimaeric heat shock gene in transgenic tobacco plants. *Transgenic Res.* **2**, 93–100.

Schuch, W., Bird, C.R., Ray, J., *et al.* (1989) Control and manipulation of gene expression during tomato fruit ripening. *Plant Mol. Biol.* **13**, 303–311.

Shirsat, A.H., Wilford, N. and Croy, R.R.D. (1989) Gene copy number and levels of expression in transgenic plants of a seed specific gene. *Plant Sci.* **61**, 75–80.

Sokal, R.R. and Rohlf, F.J. (1981) *Biometry, The Principles and Practice of Statistics in Biological Research*, 2nd Edn. San

Francisco: W.H. Freeman and Company.

Spofford, J.B. (1976) Position-effect variegation in Drosophila. In *The Genetics and Biology of Drosophila*, Volume 1c (Ashburner, M. and Novitski, E., eds). London: Academic Press, pp. 955–1018.

Steel, R.G.D. and Torrie, J.H. (1960) *Principles and Procedures of Statistics*. New York: McGraw-Hill Book Company, Inc.

Wang, K., Herrera-Estrella, L., Van Montagu, M. and Zambryski, P. (1984) Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from *Agrobacterium* to the plant genome. *Cell*, **38**, 455–462.

Weising, K., Schell, J. and Kahl, G. (1988) Foreign genes in plants: Transfer, structure, expression, and applications. *Ann. Rev. Genet.* **22**, 421–477.

White, J.L. and Kaper, J.M. (1989) A simple method for detection of viral satellite RNAs in small plant tissue samples. *J. Virol. Meth.* **23**, 83–94.