Assembly of Two Transgenes in an Artificial Chromatin Domain Gives Highly Coordinated Expression in Tobacco

Ludmila Mlynárová,*† Annelies Loonen,* Elzbieta Mietkiewska,*‡ Ritsert C. Jansen*,§ and Jan-Peter Nap*,†

*Plant Research International, Wageningen University and Research Center, NL-6700 AA Wageningen, The Netherlands, †Institute of Plant Genetics and Biotechnology, SK-950 07 Nitra, Slovak Republic, ‡Plant Breeding and Acclimatisation Institute, Młochoów Research Center, PL-05 832 Rozalin, Poland and §Biometris, Wageningen University and Research Center, NL-6700 AA Wageningen, The Netherlands

Manuscript received July 3, 2001
Accepted for publication December 4, 2001

ABSTRACT

The chromatin loop model predicts that genes within the same chromatin domain exhibit coordinated regulation. We here present the first direct experimental support for this model in plants. Two reporter genes, the E. coli ß-glucuronidase gene and the firefly luciferase gene, driven by different promoters, were placed between copies of the chicken lysozyme A element, a member of the matrix-associated region (MAR) group of chromatin boundary elements, and introduced in tobacco (Nicotiana tabacum). In plants carrying A elements, quantitative enzyme activities and mRNA levels of both genes show high correlations compared to control plants. The A element thus creates an artificial chromatin domain that yields coordinated expression. Surprisingly, enzyme activities correlated poorly with their respective mRNA levels. We hypothesize that this indicates the occurrence of “error pipelines” in data generation: systematic errors of a given analytical method will point in the same direction and cancel out in correlation analysis, resulting in better correlations. In combining different methods of analysis, however, such errors do not cancel out and as a result relevant correlations can be masked. Such error pipelines will have to be taken into account when different types of (e.g., whole-genome) data sets are combined in quantitative analyses.

A current model for the regulation of gene expression considers coordination in terms of functional units or modules. The concept of functional units of regulation supposes the presence of chromatin loop domains (Laemmli et al. 1992; Bode et al. 1996), delimited by sequences known as chromatin boundaries. Such boundaries create delimited chromatin loops and are thought to create topologically isolated units of gene regulation that shield and insulate genes located on the loop from cis-acting elements and other gene-repressing influences of the neighboring chromatin.

A whole-genome correlation analysis of yeast expression data indicated the existence of chromosomal domains of gene expression (Cohen et al. 2000), which further links structural domains with functional domains. Notably in mammalian systems, it is well established that multigene families, such as the ß-globin locus, are arranged in well and linear-time-regulated units. In plants, it is suggested that regulation and development may be less linear (Scheres 2000) and could follow a more fuzzy logic. It is of particular interest, therefore, to know whether the concept of functional units of regulation can be applied to plants as well. Previously we have demonstrated a significant reduction in between-transformant variation and position-independent expression of the ß-glucuronidase (GUS) reporter gene in transformed tobacco plants when the chicken lysozyme A element was placed at the borders of the T-DNA (Mlynárová et al. 1994, 1995, 1996). The reduced position effect of a single gene suggests the coincidence of structural loops with domains that reflect the functional organization of genome information. The chicken lysozyme A element is a member of the class of chromatin boundary elements known as matrix-associated regions (MARs). MAR elements are supposed to interact directly with the nuclear matrix, but the exact mechanism by which such elements achieve that influence is as yet unknown (Holmes-Davis and Comai 1998). Overall, a range of effects of MAR elements has been reported on gene expression levels and/or variation or stability in expression of transgenes in plants (Breyne et al. 1992; Allen et al. 1993, 1996, 2000; van der Geest et al. 1994; Han et al. 1997; Vain et al. 1999). Although the presence of MARs may protect against transcriptional silencing induced by cis interactions within repeated transgene arrays (Allen et al. 2000), their presence does not necessarily protect against the influence of strong silencing loci (Nap et al. 1997; Vaucheret et al. 1998). Also studies in cell lines demonstrate that not all MARs have the same mode of action, if any (Poljak et al. 1994; Dillon and Sabbattini 2000). The different effects are likely to arise from puta-
tive differences between MARs in combination with different assay systems, transformation approaches, and recipient cells.

Until now, only single genes have been evaluated in functional MAR assays in plants. So far, no regulatory unit of gene regulation has been demonstrated in plants. This would require us to show that two individual genes linked on a single MAR-delimited chromatin loop become coordinately regulated. Such an approach is taken here. We investigate the creation of a unit of gene regulation by focusing on the coordination of the expression of two reporter genes assembled in a single T-DNA. The _Escherichia coli_ GUS and the firefly luciferase (LUC) reporter genes, driven by two different promoters, were used. Quantitative analyses of gene expression demonstrate a significant correlation between the enzymatic activity levels as well as mRNA amounts of both reporter genes only in plants where the reporter genes are embedded in A-element DNA. This demonstrates the creation of an A-element-mediated unit of transcriptional regulation in plants and adds a new tool to the approaches available for achieving or studying coordinated expression of multiple genes in plants.

**MATERIALS AND METHODS**

*Plant transformation and genetic analysis:* Tobacco (_Nicotiana tabacum_ cv Petit Havana SR1) was grown and transformed with _Agrobacterium tumefaciens_ _H11001_ and _H9262_ into the _Nicotiana tabacum_ cv _Petit Havana SR1_ was grown and transformed with _Agrobacterium tumefaciens_ _H11001_ and _H9262_. Transgenic tobacco plants of interest were selfed. For each plant, 40–50 offspring seeds were analyzed for segregation into the T-DNA configuration they contain. To obtain pHGL, the HPT cassette as _BamHI-SalI_ fragment was ligated to the GUS cassette as _SalI-EcoRI_ fragment and the LUC cassette as _EcoRI-XbaI_ fragment in _BamHI-XbaI_ digested _pBinPLUS_. The two MAR-element-containing plasmids _pAHGLA_ and _pAHGLA_ were made in two steps. For _pAHGLA_, the subclone _pLGA_ was prepared in _pUCAP_ by cloning the LUC cassette _pLGA_ was prepared in _pUCAP_ by cloning the LUC cassette from subclone _pGLA_ in _pBinPLUS_. All binary plasmids were taken here. We investigate the creation of a unit of transcriptional regulation in plants and add a new tool to the approaches available for achieving or studying coordinated expression of multiple genes in plants.

**Plant transformation and genetic analysis:** Tobacco (_Nicotiana tabacum_ cv Petit Havana SR1) was grown and transformed with _Agrobacterium tumefaciens_ _H11001_ and _H9262_ harboring the binary plasmid of interest as described previously (Mlyna´rova´ et al. 1994). During transformation, a selection pressure of 50 μg/ml kanamycin was applied. Rooting on hygromycin was assayed on root-inducing medium (MS medium supplemented with 0.05 μg/ml of _IAA_) with 20 μg/ml hygromycin (Duchefa). Plant populations were named after the vector they contain. Transgenic tobacco plants of interest were selfed. For each plant, 40–50 offspring seeds were analyzed for segregation by germination without sterilization on quarter-strength MS medium supplemented with 50 μg/ml kanamycin or 20 μg/ml hygromycin as described previously (Conner et al. 1998).

**Vectors for plant transformation:** Standard procedures were used for DNA cloning and analysis (Sambrook et al. 1989). To facilitate cloning, the various DNA cassettes required for the binary vectors used in this study [MAR, GUS, LUC, hygromycin phosphotransferase (HPT)] were blunted, cloned in the standard cloning vectors _pUC18_ (New England BioLabs, Beverly, MA) or _pSK+_ (Stratagene, La Jolla, CA) and _pSK+_. The chicken lysozyme _A_ element was isolated as a _BamHI-XbaI_ fragment from plasmid _pUC-B1-X1_ (PhI-Van and Strätling 1988), blunted, and cloned into the _Smal_ site of _pUC8_. The GUS gene fused to the potato light-regulated _Lhca3-St.1_ promoter and _nos_ terminator was isolated as an _XbaI-EcoRI_ fragment from plasmid _pPPG_ (Nap et al. 1993a). This GUS cassette was blunted and cloned in the _EcoRV_ site of _pSK+_. The _A_ promoter for the cytotoxic form of firefly luciferase was obtained from plasmid _pSP-luc+_(Promega, Madison, WI) and supplied with the single 35S CaMV promoter with an N-terminal SV40 nuclear localization signal (NLS; van der Kroel and Chua 1991) and the _nos_ terminator. The presence of the NLS was shown to have no effect on the activity of the luciferase gene (van Leeuwen et al. 2000). The _LUC_ cassette was isolated as a _BamHI-SalI_ fragment, blunted, and cloned in the _EcoRV_ site of _pSK+_. The hygromycin phosphotransferase gene driven by the _nos_ promoter and carrying the polyadenylation signal _pAg4_ was obtained from plasmid _pCVC720_ (Koncz et al. 1994). This HPT cassette was isolated as a _HindIII-Asp718_ fragment, blunted, and cloned in the _EcoRV_ site of _pSK+_. All plant transformation vectors were prepared in the binary vector _pBinPLUS_ (van Engelen et al. 1995), which is a _pBIN19_ derivative with the _nos_ promoter-driven kanamycin resistance selectable marker gene. Compared to _pBIN19_, _pBinPLUS_ also has additional _AscI_ and _Pad1_ cloning sites. The _BinPLUS_ multiple cloning sites are also present in the _pUC19_-based plasmid _pUCAP_ (van Engelen et al. 1995). The binary vectors prepared are named according to the T-DNA configuration they contain. To obtain _pHGL_, the _HPT_ cassette as _BamHI-SalI_ fragment was ligated to the GUS cassette as _SalI-EcoRI_ fragment and the LUC cassette as _EcoRI-XbaI_ fragment in _BamHI-XbaI_-digested _pBinPLUS_. The two MAR-element-containing plasmids _pAHGLA_ and _pAHGLA_ were made in two steps. For _pAHGLA_, the subclone _pLGA_ was prepared in _pUCAP_ by cloning the LUC cassette as _SalI-EcoRI_ fragment, the _GUS_ cassette as _EcoRI-KpnI_ fragment, and the _MAR_ cassette as _KpnI-XbaI_ fragment. The second subclone _pAH_ was made by cloning the _MAR_ cassette as _KpnI-BamHI_ and the _HPT_ cassette as _BamHI-SalI_ fragment in _pUCAP_. Subsequently, the binary vector _pAHGLA_ was obtained by ligation of the _KpnI-AscI_ fragment from the _pAH_ subclone and _AscI-XbaI_ fragment from the _pGLA_ subclone in _pBinPLUS_. A similar strategy was used to obtain _pAHGLA_. Subclone _pGLA_ was prepared in _pUCAP_ by cloning the _GUS_ cassette as a _SalI-EcoRI_ fragment, the _LUC_ cassette as _EcoRI-KpnI_ fragment, and the _MAR_ cassette as _KpnI-XbaI_ fragment. The _pAHGLA_ binary vector was obtained by cloning the _KpnI-AscI_ fragment from subclone _pAH_ and the _AscI-XbaI_ fragment from subclone _pGLA_ in _pBinPLUS_. All binary plasmids were conjugated to _A. tumefaciens_ _LBA4404_ and verified by reintroduction into _E. coli_ as previously described (Mlyna´rova´ et al. 1994). The structure of the T-DNA configurations after integration into the plant genome is shown in Figure 1. In designing these three T-DNA vectors, care was taken that (i) the selectable marker cassette neomycin phosphotransferase (NPT) is in all cases subject to similar random influences of surrounding plant DNA and (ii) no directly neighboring genes were transcribed in opposite directions. Although an obvious consequence of these requirements is that not all genes have precisely the same relative orientation with respect to the _Agrobacterium_ T-DNA border sequences, the gene configurations resulting from integration (Figure 1) allow optimal assessment of coordinated gene expression in planta.

**Determination of GUS and LUC enzymatic activity:** For GUS and LUC assays, plant extracts were prepared by grinding 0.8-cm² leaf discs of greenhouse-grown plants in luciferase lysis buffer (100 mM potassium phosphate, pH 7.8; 9.2% Triton X-100; 1 mM dithiothreitol) according to the manual of the luciferase assay kit (Tropix, Bedford, MA) and cleared by centrifugation for 5 min at 4°. GUS activity was determined as described previously (Nap et al. 1999) and expressed as picomoles of methyl-umbelliferyl per minute per microgram of soluble protein. Luciferase activity was determined in the same leaf extracts using a luciferase assay kit (Tropix) according to the recommendations of the manufacturer. Bioluminescence reactions were performed in white microtiter plates (2FB; Dynatech, Chantilly, VA) with 10 μl of plant extract and 10 μl of the luciferase assay buffer, using a Labsystems (Marlboro, MA) Luminoskan DS dual injector luminometer (van Leeuwen et al. 2000). The luciferase reaction was started by automatic injection of 20 μl of the substrate buffer luciferin (Tropix). Five seconds after injection the light produced during 10 sec of reaction was measured as luciferase activity. Luciferase activity was subsequently expressed as relative light units (RLU) per milligram of soluble protein. A dilution of purified luciferase (Roche, Indianapolis) was used
Coordinated Gene Expression in Plants

**Figure 1.**—Structure of the T-DNA region present in plants. In all three cases, the two selection genes, the hygromycin phosphotransferase gene (HPT) and the neomycin phosphotransferase gene (NPT), are driven by the nos promoter. The arrow gives the direction of transcription of the selection genes. The GUS gene is driven by the potato *Lhca3* St.1 promoter (*Lhca3*) and the luciferase gene (LUC) by the single 35S CaMV promoter (CaMV). The arrows give the direction of transcription by these two promoters. The chicken lysozyme matrix-associated region known as A element is represented by open triangles labeled A. The arrows above the A element give the relative orientation of the A-element DNA. These DNAs are not drawn to scale.

Restriction enzymes sites used for DNA blot analysis and the expected (minimum) sizes of the relevant hybridizing fragments in plant genomic DNA are indicated. The remains of the T-DNA border sequences are indicated by solid (left border) and open (right border) triangles; curved boxes indicate flanking genomic DNA and dashed arrows indicate hypothetical restriction sites in the flanking plant DNA.

Data were evaluated using the analysis tools of both Excel 97 and GenStat5 (Genstat 5 Committee 1993).

**DNA and RNA analysis:** Genomic DNA was isolated from tobacco leaf material by the urea-phenol extraction procedure as described previously (Mlynárová et al. 1994). Genomic DNA was digested with the appropriate restriction enzymes *VspI* and *HindIII* (Life Technologies) and *AscI* (New England BioLabs) using the conditions recommended by the manufacturer, separated on 0.8% agarose gel and electrophoreted on Hybond-N+ membrane (Amersham, Buckinghamshire, UK). Probes were isolated as restriction fragments from agarose gel and were labeled with [32P]dATP using a Megaprime DNA labeling system (Amersham). Prehybridization was performed at 65°C in hybridization solution (10% dextran sulfate, 1% SDS, 1 M NaCl, 0.1 mg/ml sonicated salmon sperm DNA) for at least 1 hr. Hybridization was performed overnight at 65°C in a hybaid hybridization oven (Hybaid). After hybridization, the membrane was washed first in 2× SSC, 1% SDS, followed by washing in 1× SSC, 1% SDS at 65°C. Hybridizing signals were visualized by autoradiography on Fuji 100NF films and quantified using a Bas2000 PhosphorImager (Fuji) with BasReader and TINA software (Raytest). For rehybridization, blots were stripped by four to five periods of short incubation (2 min) in a boiling solution of 0.01% SDS and 0.01× SSC and checked for the absence of any remaining signal. Hybridizing signals were visualized by autoradiography on Fuji 100NF films and quantified using a Bas2000 PhosphorImager (Fuji) with BasReader and TINA software (Raytest). Statistical analyses were performed with the analysis tools of Excel 97 as well as with GenStat5 (Genstat 5 Committee 1993).

**RESULTS**

Generation and molecular characterization of tobacco transformants: The T-DNA configurations introduced in tobacco plants are shown in Figure 1. For
TABLE 1
Characteristics of T-DNA integrations in the three populations of plants

<table>
<thead>
<tr>
<th>Integration*</th>
<th>Population name and no. of plants in the population*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HGL</td>
</tr>
<tr>
<td>Total (%)</td>
<td>37  (100)</td>
</tr>
<tr>
<td>Complex (%)</td>
<td>28  (76)</td>
</tr>
<tr>
<td>Simple (%)</td>
<td>9   (24)</td>
</tr>
<tr>
<td>One-copy (%)</td>
<td>5   (14)</td>
</tr>
<tr>
<td>Two-copy (%)</td>
<td>3   (8)</td>
</tr>
<tr>
<td>More-than-two copies (%)</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

* Plant populations are named after the T-DNA they carry (see Figure 1).
* Type of T-DNA integration as determined by DNA blot analysis.
* Total population of plants with at least one complete integration of the T-DNA.
* Population of plants with a complex integration of the T-DNA (based on an unequal number of left and right border fragments).
* The percentage of plants relative to the corresponding total population is given in parentheses.
* Population of plants with a simple integration of the T-DNA, based on the equal number of left and right border fragments.
* Number of plants with one, two, or more than two simple T-DNA copies integrated intact.

In Table 2, the descriptive statistics are given for these populations.
six data sets. Comparing the mean activity values, these data suggest that the gene present in the middle of the MAR-delimited loop tends to have a higher activity than when the same gene is closer to the A element. The AHGLA population, with the GUS gene in the middle, has an average GUS activity of 278.6 pmol methylumbelliferone per minute per microgram of protein, whereas the AHLGA population has an activity of only 148.4, almost half that of the AHGLA population. For the LUC gene this trend is reversed.

Comparing variances, the MAR-containing populations show a significantly reduced variation in GUS activities compared to the control population HGL. The AHGLA and AHGLA populations show 4.3- and 3.5-fold reduction of variance compared to HGL, which is highly significant (Table 2). As shown before (Mlynárová et al. 1994, 1995), the presence of the A element is accompanied by a significant reduction of the variation of GUS activity between independent transformed plants. In contrast, no significant differences in variation were observed for the LUC activity data compared to the control population (Table 2), suggesting that the MAR elements have had no detectable influence on the variation in LUC activity. Further analyses (see below) show that this result is due partly to the presence of complex integrations in this total population.

Of special interest in this study is the relation between the activities of the two genes present on the composite T-DNA. Figure 2, G–I, shows the double logarithmic plot of GUS activity for each transformant plotted against the LUC activity of the same transformant. Table 3 gives the product moment correlation coefficients (R) for these associations. In the control population HGL, there is no apparent correlation between the activities of both genes (R = 0.24), whereas the relationship between both activities is markedly increased in the AHGLA (R = 0.83) and AHLGA (R = 0.81) populations. Spearman’s rank correlation, which is not dependent on the logarithmic transformation procedure, yields the same conclusions (not shown). These results indicate that for the MAR-containing populations, plants with high GUS activity also show high LUC activity and vice versa. Part of the improved correlation in the MAR-containing populations could be due to a copy number effect. To remove such a copy number effect from the plot, the ratio of logarithmic GUS and LUC activity [i.e., ln(GUS)/ln(LUC)] is plotted in Figure 2, J–L. The nearly straight line obtained in the case of the AHGLA and AHGLA populations shows that GUS and LUC activity in all the MAR-containing plants yields virtually the same ratio, whereas the ratio in the HGL control population is much more variable. These data are quantified and analyzed in Table 4. The data in Table 4 illustrate that the variance of the ln(GUS)/ln(LUC) ratio of the AHGLA and AHGLA populations is highly significantly reduced compared to the control population HGL. The activity levels of the two genes become more associated, supposedly due to a common cause, i.e., the presence of the MAR elements.

Plants with complex integrations may show post-transcriptional gene silencing (Muskens et al. 2000) and tend to have more unstable and less predictable transgene expression levels. Such plants can influence the statistical evaluation of populations considerably. In Table 2, the descriptive statistics of all plants classified as complex are given. In Figure 2, A–L, the complex plants are plotted as open circles. Comparing the total populations with the complex subsets, it can be concluded that there is little difference (Tables 2–4). On the basis of correlation coefficients (Table 3) as well as on ln(GUS)/ln(LUC) variances (Table 4), the two MAR-containing populations behave more consistently than the complex subpopulation from the HGL control.

Analyses of GUS and LUC enzymatic activities in plants with simple T-DNA integrations: To eliminate the potential influence of complex T-DNA integrations, the quantitative analyses were restricted to those plants that had simple integration patterns. This reduces the total number of plants available for analysis to 42. Table 2 gives the descriptive statistics of the data sets of these plants. In Figure 2, the simple plants are indicated by the sum of the red and the blue circles. In Figure 2, A–C, the GUS activity data are plotted and in Figure 2, D–F, the LUC activity data, respectively. Figure 2, G–I, shows the double logarithmic plot of GUS activity against the LUC activity and Figure 2, J–L, the copy-number-corrected ln(GUS)/ln(LUC) ratio. Table 3 gives the GUS/LUC correlation and Table 4 gives the mean and variance of the ratios plotted in Figure 2, J–L. For the GUS activity, the two MAR-containing populations still show a significantly reduced variation compared to the control population HGL. Removing the complex integrations from the analysis now reveals a 4.8-fold significant reduction in variance for LUC activity in the AHGLA population compared to the control population HGL (Table 2). The variance in the AHGLA luciferase activity is ~2-fold lower than the variance in the HGL population, which is barely not significant (at P = 0.05; Table 2) due to copy number effects (Table 4). When copy number effects are eliminated (Table 4), the AHGLA simple population also shows a significant reduction in the variation of LUC activity. The correlation coefficients (Table 3) follow the trends given by the total population: in the control population HGL, there is still no correlation between the activities of both genes (R = 0.31), whereas the relationship is obvious in the MAR-containing AHGLA (R = 0.80) and AHLGA (R = 0.73) populations.

In plants with multiple simple integrations of transgenes, these transgenes may (epigenetically) interact and deviate from the expected copy-number-dependent additivity of gene expression. Therefore, a final analysis was based on all plants with a simple one-copy integration of the T-DNA. The total number of plants available for this analysis is 24. Table 2 gives the descriptive statis-
Figure 2.—Quantified enzymatic activities in all transformed plants in the three populations. The population name is given at the top. (A–C) The ln of the GUS activity. (D–F) The natural logarithm of the luciferase activity. (G–I) The double logarithmic plot of GUS activity plotted against the LUC activity of the same transformant. (J–L) The ratio of the logarithmic GUS and LUC activity. In A–L, plants with a one-copy integration of the T-DNA are indicated in blue, plants with two or more simple T-DNA integrations are in red, and all other plants with complex integrations are represented by open circles. Panels related to the GUS gene are in blue, and those related to the LUC gene are in yellow. Panels combining the two genes are in a combination of both colors.
Coordinated Gene Expression in Plants

### TABLE 2

<table>
<thead>
<tr>
<th>Integration</th>
<th>Population</th>
<th>GUS activity</th>
<th>LUC activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Mean</td>
<td>Var.</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGL</td>
<td>37</td>
<td>5.01</td>
<td>1.43</td>
</tr>
<tr>
<td>AHLGA</td>
<td>27</td>
<td>5.00</td>
<td>0.33</td>
</tr>
<tr>
<td>AHGLA</td>
<td>33</td>
<td>5.63</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>Complex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGL</td>
<td>28</td>
<td>5.24</td>
<td>1.33</td>
</tr>
<tr>
<td>AHLGA</td>
<td>12</td>
<td>4.96</td>
<td>0.57</td>
</tr>
<tr>
<td>AHGLA</td>
<td>15</td>
<td>6.02</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Simple</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGL</td>
<td>9</td>
<td>4.32</td>
<td>1.20</td>
</tr>
<tr>
<td>AHLGA</td>
<td>15</td>
<td>5.02</td>
<td>0.16</td>
</tr>
<tr>
<td>AHGLA</td>
<td>18</td>
<td>5.30</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>One-copy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGL</td>
<td>5</td>
<td>3.64</td>
<td>1.06</td>
</tr>
<tr>
<td>AHLGA</td>
<td>10</td>
<td>4.69</td>
<td>0.08</td>
</tr>
<tr>
<td>AHGLA</td>
<td>9</td>
<td>4.75</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Two-copy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGL</td>
<td>3</td>
<td>5.10</td>
<td>0.12</td>
</tr>
<tr>
<td>AHLGA</td>
<td>4</td>
<td>5.45</td>
<td>0.06</td>
</tr>
<tr>
<td>AHGLA</td>
<td>9</td>
<td>5.73</td>
<td>0.12</td>
</tr>
</tbody>
</table>

$^a$Type of T-DNA integration as determined by DNA blot analysis (see text and Table 1); total is complex plus simple; simple is further subdivided into one-copy and two-copy plants.

$^b$Populations are named according to the T-DNA construct the plants contain (see Figure 1).

$^c$Number of plants in the population.

$^d$Means are based on the natural logarithms of enzyme activities.

$^e$Variance.

$^f$Fold reduction of variance with respect to the corresponding control population HGL.

$^g$One-tailed probability according to the $F$ test for homogeneity of variances in comparison to the corresponding control population HGL.

$^h$(—) Control population HGL.

$^i$Not applicable.

$^j$Indication of the probability value at the 95% confidence level; ***, significant at $P < 0.001$; **, significant at $P < 0.01$; *, significant at $P < 0.05$; NS~-, barely not significant ($0.05 < P < 0.1$); NS, not significant (at $P > 0.1$).

In Figure 2, the one-copy populations are indicated by the blue circles. In Figure 2, the GUS activity data are plotted in A–C and the LUC activity data are plotted in D–F. In Figure 2, G–I, the double logarithmic plot of GUS activity against the LUC activity is given and in Figure 2, J–L, the derived ln(GUS)/ln(LUC) ratio is given. Table 4 gives the mean and variance of the ratios plotted in Figure 2, J–L. For both GUS and LUC the variation in one-copy AHGLA and AHLGA plants is reduced significantly ($P < 0.001$), up to 18-fold, compared to the HGL population (Table 2), as is also apparent from Figure 2, A–C (for GUS) and Figure 2, D–F (for LUC). Both the one-copy AHGLA and the AHLGA populations cluster together in double logarithmic plots (Figure 2, H and I, blue circles). In contrast, the HGL data show a considerable scatter (Figure 2G, blue circles), despite the low number of only 5 plants present in this population. The correlation coefficient is not a very suitable way of demonstrating a clustering of plants toward a single point. Due to the small population sizes, none are significantly different from zero (data not shown). In the case of the non-MAR HGL one-copy plants, the covariance of GUS and LUC enzymatic activity is much higher than for the MAR-containing one-copy plants (data not shown). The difference between the HGL, AHGLA, and AHLGA one-copy populations is also characterized by a significant reduction in the variance of the ln(GUS)/ln(LUC) ratio (Figure 2, J–L; Table 4). This shows the highly coordinated expression of the GUS and LUC activities in the MAR-containing one-copy populations.

### Relationship between GUS and LUC enzymatic activities and mRNA levels

The data demonstrating the efficacy of the chicken lysozyme A element in the reduction of position effects are based on enzymatic activities. This implicitly assumes linear relationships between transcription rate, steady-state mRNA accumulation, amount of protein, and the enzymatic activity of that protein. One of the concerns is whether and to what extent the quantitative activity data reflect the steady-state RNA accumulation of the genes used. Therefore, the analyses were extended to the RNA levels of GUS and LUC in the three populations of transgenic plants. The remainder of almost every leaf from which a sample was taken...
TABLE 3
Correlation of GUS and luciferase expression data

<table>
<thead>
<tr>
<th>Integration</th>
<th>Population</th>
<th>GUS activity-</th>
<th>GUS activity-</th>
<th>LUC activity-</th>
<th>GUS mRNA-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LUC activity</td>
<td>GUS mRNA</td>
<td>LUC mRNA</td>
<td>LUC mRNA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>HGL</td>
<td>37</td>
<td>0.24</td>
<td>NA</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33</td>
<td>0.55</td>
<td>NA</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>AHGLA</td>
<td>27</td>
<td>0.81</td>
<td>***,</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27</td>
<td>0.75</td>
<td>NS~</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>AHGLA</td>
<td>33</td>
<td>0.83</td>
<td>***</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.80</td>
<td>*</td>
<td>30</td>
</tr>
<tr>
<td>Complex</td>
<td>HGL</td>
<td>28</td>
<td>0.23</td>
<td>NA</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>0.57</td>
<td>NA</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>AHGLA</td>
<td>12</td>
<td>0.86</td>
<td>***</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.85</td>
<td>NS~</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>AHGLA</td>
<td>15</td>
<td>0.79</td>
<td>***</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>0.82</td>
<td>NS~</td>
<td>14</td>
</tr>
<tr>
<td>Simple</td>
<td>HGL</td>
<td>9</td>
<td>0.31</td>
<td>NA</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>0.56</td>
<td>NA</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>AHGLA</td>
<td>15</td>
<td>0.73</td>
<td>**</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>0.66</td>
<td>NS</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>AHGLA</td>
<td>18</td>
<td>0.80</td>
<td>**</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>0.67</td>
<td>NS</td>
<td>16</td>
</tr>
</tbody>
</table>

* Comparisons are based on the natural logarithms of the data on enzymatic activity and mRNA amount.
* Type of T-DNA integration as determined by DNA blot analysis (see text and Table 1); total is complex plus simple.
* Populations are named according to the T-DNA construct that the plants contain (see Figure 1).
* Number of combined data points.
* Product-moment (i.e., the common) correlation coefficient corrected for the bias in the case of small samples from bivariate normal distributions (Sokal and Rohlf 1995). The correlation coefficients given in italics are not significantly different from zero at P = 0.05 according to the t test for significance.
* One-tailed probability according to the Fisher z transformation with z approximation (Zar 1996) for equality of the correlation coefficient in comparison to the corresponding control population HGL.
* Not applicable.
* Indication of the probability value at the 95% confidence level; ***, significant at P ≤ 0.001; **, significant at P ≤ 0.01; *, significant at P ≤ 0.05; NS~, barely not significant (0.05 < P ≤ 0.1); NS, not significant (at P > 0.1).

for the determination of the enzymatic GUS and LUC activities was used for total RNA isolation. The relative amounts of GUS and LUC transcripts as well as 18S rRNA amounts were determined with radioactively labeled probes on RNA blots. In Figure 3, representative examples of the RNA blot analyses for each of the three populations are shown. Data were quantified with the help of a PhosphorImager. The GUS and LUC RNA signals were normalized for the amount of RNA loaded relative to the rRNA signal. Also for the normalized RNA amounts, a natural logarithmic transformation was appropriate for statistical evaluation (analysis not shown), so all subsequent analyses were performed after such a transformation.

In Figure 4, the analysis of the RNA data for the total populations is shown. As in Figure 2, this is the sum of the red, blue, and open circles. The relationship between GUS activity and amount of GUS mRNA is plotted in Figure 4, A–C, and the relationship between LUC activity and amount of LUC mRNA is plotted in Figure 4, D–F. The corresponding correlations are given in Table 3. Whereas there is no correlation between LUC mRNA amount and LUC activity in the HGL population (R = 0.07), the equivalent correlation in the MAR-containing plants is still poor but markedly improved (R = 0.54/0.53) and significantly different from zero. For GUS, the same trend is apparent: the HGL population shows a rather poor correlation (R = 0.55), whereas the AHGLA and AHGLA populations are improved (R = 0.75/0.80). These data indicate, therefore, that the overall correlation between enzymatic activity and mRNA amount is disappointing. The presence of the MAR elements helps to improve that correlation. In Figure 4, G–I, the GUS mRNA is plotted against the LUC mRNA amount; the corresponding correlation coefficients are given in Table 3. Despite the poor correlations between enzymatic activity and mRNA, the correlations between GUS and LUC mRNA are high. As shown above for the activity data, for the mRNA amount data a major part of the correlation also is likely to be due to copy number effects. Figure 4, J–L, shows the plot of the ln(GUS mRNA)/ln(LUC mRNA) ratio and the accompanying statistics are given in Table 4. Compared to the control population HGL, a reduction in the ln(GUS mRNA)/ln(LUC mRNA) ratio is observed. Therefore, the expression of both genes has become coordinated on the level of steady-state RNA amount as well. Similar analyses of the subset of plants with simple copies (sum of red and blue circles, Figure 4, A–L; Tables 3 and 4) and the simple one-copy plants (blue circles, Figure 4, A–L) further substantiate the results obtained for the total population. Due to the relatively small population sizes, the confidence limits of the statistical parameters estimated are large, so only trends can
be pointed out. The relationship between GUS mRNA and GUS activity, on the one hand, and LUC mRNA and LUC activity, on the other hand, improve in the two MAR-containing populations compared to the HGL control.

**DISCUSSION**

The expression of two different genes placed between chicken lysozyme A-element DNA becomes highly correlated and coordinated, in contrast to the situation where the genes are flanked by only selectable marker genes. Both enzymatic activity levels and steady-state RNA levels of the two genes become coordinated, suggesting that the influence of the A elements is on the level of transcription. The artificial chromatin domain created by the chicken lysozyme A element is apparently establishing a functional unit of gene regulation in plants.

Whereas the phenotypic effects observed in the presence of the A elements are obvious, the molecular mechanism of this biological activity of the elements is a controversial issue. It is an open question whether the observed coordination is due to nuclear matrix binding or to some other specific, but yet unknown, characteristic of the A element. The presence of the A element on transforming DNA could, for example, somehow help to target the transgenes to active regions of the genome (Allen et al. 2000). In different laboratories, different results have been obtained with different MAR elements, genes, and recipient cells or species, suggesting that results cannot easily be generalized. Our control (see Figure 1) is devised in such a way that the two selectable marker genes (NPT and HPT) surround the reporter genes (GUS and LUC). Both marker genes are active in the plants analyzed. These genes are evaluated as an alternative to the A element in shielding transgenes from the influences of neighboring chromatin. Bhattacharyya et al. (1994) showed a reduced GUS variability using a T-DNA with two selectable markers located at the T-DNA borders, similar to the transgene configuration analyzed here. Compared to previous results (Mlynárová et al. 1994), the observed reduction in variability is not merely a matter of selectable marker DNA around the gene(s) of interest. A reduction in variability requires the presence of DNA with the characteristics of the A element. Besides its presumed role in nuclear architecture, the A element could collaborate with enhancer and/or promoter sequences (Forrester et al. 1994), generate an extended domain of open chromatin (Zhao et al. 1993; Jenuwein et al. 1997), or influence transcription and pre-mRNA processing components (Göhring and Fackelmayer 1997; Nayler et al. 1998).

Another issue of debate is the role of the configura-
variation of LUC activity could be demonstrated only in the subpopulations of one-copy plants (Table 2). No differences in variance of LUC enzymatic activities were observed between the MAR-containing population and the HGL control population. A possible explanation is that the CaMV promoter driving the LUC gene is more prone to epigenetic effects than the Lhca3.St.1 promoter driving the GUS gene (Mlynárová et al. 1996). In one-copy plants, the MAR elements reduce the position effect in the luciferase activity as much as they reduce the position effect of GUS (GUS: 13- and 18-fold reduction; LUC: 9- and 15-fold reduction). Therefore, in this subpopulation the MAR elements also do protect the luciferase gene from the influences of the surrounding chromatin and this reduction of position effects reveals the correlation between the activity levels.

In subsets of the populations defined on the basis of T-DNA configuration, the differences in variance between MAR and non-MAR plants become larger. This confirms that complex integrations and copy number dependence of gene activity influence the variance in the populations of transgenic plants. Complex integrations are often associated with gene-silencing phenomena (Muskens et al. 2000). Comparisons of the three complex populations generated here show that the presence of MAR elements improves the expression characteristics in terms of variance reductions and correlations (Tables 2–4). The presence of MAR elements may therefore protect (in part) against transcriptional silencing induced by cis interactions within complex and/or repeated transgene arrays (Allen et al. 2000). A copy number effect resulting in additive enzyme activities of multiple integrations obviously increases the variance when the total population is evaluated. The moment populations differ in the relative proportion of complex and multicopy integration events, comparisons and conclusions about MAR action may be erroneous or uninformative. The differences between the GUS and LUC data in the total populations are likely to be due partly to the characteristics of the LUC protein. Obviously, both mRNA and protein stability play an important role in establishing both the level and the kinetics of gene expression. The LUC protein is degraded rapidly, notably in the absence of its substrate (Thompson et al. 1991). As a result, it provides a better and instantaneous monitor of gene expression, but will also be more variable than a long(er)-living reporter molecule such as GUS.

The data show an unexpected and surprisingly poor correlation between enzymatic activity and RNA amount (Table 3) for both GUS and LUC. In large-scale comparisons in yeast of the transcriptome with the proteome, only the high abundant proteins showed a good correlation between mRNA amount and protein. Out of 106 genes analyzed, no less than 95 had a correlation of 0.5 or less, the majority showing a correlation of <0.3 (Gygi et al. 1999). Generally, lack of correlation between RNA and protein levels is thought to be due to post-transla-

**Figure 3.**—Autoradiograph of a RNA blot showing the RNA amounts in a random subset of plants from each of the three populations. Total RNA was separated on formaldehyde gels, blotted, probed successively with labeled probes for the GUS gene, the LUC gene, and the pea 18S ribosomal sequences (rib), and visualized by phosphorimaging. The phosphor images shown have had different exposure times that were selected to give approximately comparable signals.

The variation of genes within the A-element-delimited DNA loop. The two test constructs evaluated here (see Figure 1) were devised in such a way that they do not have an A element immediately upstream of the promoter of either the GUS or the LUC reporter gene. This was based on our previous results (Mlynárová et al. 1994, 1995) and follows suggestions that transcription may be hampered by too close an association with the nuclear matrix (Razin et al. 1995; Strätling and Yu 1999). Moreover, the configuration of a nos promoter relative to the dCaMV and Lhea3.St.1 promoter-gene fusions is kept the same to eliminate potential problems with any putative readthrough. The analyses show that the LUC gene yields most activity in the AHLGA population when in the middle of the chromatin loop. The GUS gene shows the same trend. When in the middle of the presumed loop of A-element-delimited DNA, the gene tends to be most active. This indicates that a gene may be most accessible for the transcription machinery when placed away from the A elements.

The variation in GUS activity in the MAR-containing populations analyzed here is significantly lower than in the non-MAR control analyzed previously (Mlynárová et al. 1994). Therefore, the MAR elements also reduce the position effect when the presumed DNA loop between the two elements is enlarged. A reduction in the
Figure 4.—Quantified GUS and LUC mRNA amounts in all transformed plants in the three populations. Total RNA was analyzed as indicated in Figure 3. The relative amounts of RNA transcripts were quantified with a PhosphorImager; the GUS and LUC signals were normalized for the amount of RNA loaded relative to the hybridization with the ribosomal probe. (A–C) The double logarithmic plot of GUS activity and normalized GUS mRNA. (D–F) The natural logarithm of LUC activity and normalized LUC mRNA. (G–I) The double logarithmic plot of the amount of GUS mRNA plotted against the LUC mRNA in the same transformant. (J–L) The ratio of the logarithmic GUS and LUC mRNA amounts. In all panels, plants with a one-copy integration of the T-DNA are indicated in blue, plants with two or more simple T-DNA integrations are in red, and all other plants with complex integrations are represented with open circles. Panels related to the GUS gene are in blue, and those related to the LUC gene are in yellow. Panels combining the two genes are in a combination of both colors.
tional modifications and/or relative differences between mRNA and protein turnover. The latter may explain in part the difference between the GUS and LUC data. The improved correlation upon the presence of MAR elements suggests that the positive influence of the MAR elements extends beyond their role in the nucleus. They may somehow influence the export from the nucleus, RNA turnover, or possibly even the translational efficiency of the mRNA in the cytoplasm. The relatively poor correlation between RNA and enzyme activity levels creates the following situation: for the MAR-containing populations, GUS and LUC enzymatic activities are well correlated and GUS and LUC mRNA levels are well correlated, but GUS and notably LUC activities are not well correlated with their respective mRNA levels. At first, this is somewhat counterintuitive. A possible statistical explanation is the following: the experimental procedures of protein isolation and kinetic analyses, on the one hand, and RNA analysis, on the other hand, each involve a number of experimental steps, each with its own errors. Some of these errors will not be completely independent from one sample to the other, but rather work in the same direction. For example, the same isolate and protein determination is used for both the GUS and the LUC kinetic measurements. Similar arguments hold for the determination of RNA levels. When performing correlation analyses (Table 3) or when taking the activity ratio as the parameter of analysis, such as in Figure 2, J–L, and in Figure 4, J–L, all method-dependent errors cancel out and the result will be more accurate, and therefore a more relevant, estimate. However, when enzymatic activities are combined with RNA levels, the methodological errors become much more independent. They will not cancel out, but in contrast give rise to more “noise.” As a result of this increased noise, correlations can become disguised and/or undetectable. For the data presented here, the consequence of such “error pipelines” is that the correlations between enzymatic activity levels, on the one hand, and RNA levels, on the other hand, both showing coordinated expression of the genes investigated, are more informative than the poor correlation between enzymatic activity and RNA level. The concept of data generation-dependent error pipelines would seem to need more attention in the bioinformatics of the analyses of different types of data, e.g., in whole-genome data sets.

Over the years, there have been numerous investigations into the possibilities and mechanisms of coordinated expression in plants. Dean et al. (1988a,b) showed a reasonable coordinated expression, quantified as RNA, of the chloramphenicol acetyl transferase and octopine synthase genes, when flanked with large similar, but not identical, genomic regions of well-expressed petunia rbcS gene regions and rbcS promoters. This may have been due to the presence of MAR-like sequences in those regions of petunia DNA. Other attempts with multiple genes have been less successful. Gidoni et al. (1988) could achieve the coordinated transgene expression only by maintaining the normal linkage of petunia divergently expressing Cab22L and Cab22R genes. The divergent mas1,2 promoter did not yield a correlated expression of two genes (Peach and Velten 1991), nor did the simultaneous introduction of two genes on a single plasmid into tobacco by particle bombardment (Leech et al. 1998). For complex metabolic or regulatory pathways, multiple genes are generally required to study or achieve the desired alteration(s). In a transgenic setup, this requires sequential genetic modification (Krohn et al. 1998), stacking by successive crossing (Porrier et al. 1992, 2000; Nawrath et al. 1994; Ma et al. 1995), assembling the genes of interest on a single transforming molecule (van Engelen et al. 1994; Jach et al. 1995; Leech et al. 1998), or cotransformation with multiple plasmids (Chen et al. 1998; Maqbool and Christou 1999). Results show that when genes are combined, expression levels are highly variable and, when analyzed, there is poor correlation of the expression of cotransduced genes (Jach et al. 1995; Leech et al. 1998; Maqbool and Christou 1999; Ye et al. 2000). Alternative approaches using a bicistronic transcriptional unit (Iida et al. 1992; Lough et al. 1997), a self-cleaving polyprotein (Marcos and Beachy 1994, 1997; Dasgupta et al. 1998; Halpin et al. 1999), or a bifunctional protein (Elmayan and Tepfer 1994) seem more specific and less versatile than the chromatin-structure-based approach demonstrated here. For any strategy and for each gene, the best performing plant must be selected, characterized, proven sufficiently stable over time and over generations, and be compatible with the gene(s) with which it must be combined. The results presented here indicate that the use of a chromatin boundary element such as the chicken lysozyme A element may contribute to establishing a coordinated, high-level, and stable expression of multiple introduced genes. Notably the coordinated expression of two different promoter-gene combinations may have useful applications. It may allow the creation of a reference gene that is easily assayed for and functions as a “pointer” or internal standard for the transgene of interest that cannot be assayed for so easily. For example, the analysis of a gene desired to perform in the fruit of a woody plant species that takes several years to set its first fruit now requires the long waiting for that first fruit. With a proper and coordinately expressed reference gene already in the seedling stage, a selection can be made of trees that have a considerably higher likelihood of the gene of interest doing well in the fruit. The overall, significantly higher predictability of gene action when embedded in A-element DNA would make such strategies feasible and attractive.

The authors thank Alexander van der Kroel and Wessel van Leeuwen (Plant Physiology Wageningen UR) for the LUC gene and fruitful discussions on the topic, Andy Pereira for plasmid pPCV729, Arjen et al.
Jonas (Wageningen UR, Toxicology) for indispensable help with the in vitro luciferase measurements, the Plant Research International Greenhouse team for plant care, and Lonneke van der Geest for useful suggestions to improve the manuscript. The research was supported by the Life Sciences Foundation (SLW), which is subsidized by the Netherlands Organization for Scientific Research (NWO) as well as by program subsidy 347 from the Dutch Ministry of Agriculture, Nature Management and Fisheries.

LITERATURE CITED


Communicating editor: O. Savolainen