ABSTRACT

HALWEG, CHRISTOPHER JAY. The Use of Flow Cytometry to Investigate the Effects of Matrix Attachment Regions on Transgene Expression in Plant Cells. (Under the direction of Dr. Steven Spiker and Dr. William Thompson)

Many studies in both plant and animal systems have shown that Matrix Attachment Regions (MARs) can increase expression of transgenes in whole organisms or cells in culture. MARs are AT-rich sequences of DNA that bind *in vitro* to the proteinacous filament-like structure within the nucleus called the nuclear matrix. In our investigation of transgenic *Nicotiana tabacum* NT-1 cells in culture, we have observed that transgene expression is often variegated. In other words, some cells in an isogenic population do not express the transgene, and/or other cells within the same population express the transgene at varying levels. The question was raised: Do MARs increase transgene expression by altering variegation? More specifically, do MARs increase the percentage of cells expressing the transgene, increase the magnitude of expression in expressing cells, or both?

In order to address these questions, it was necessary to quantitate transgene expression variegation at the resolution of individual cells. We chose to measure Green Fluorescent Protein (GFP) expression in individual tobacco NT-1 cells by flow cytometry. In order to analyze individual cells and because NT-1 cells in culture grow as filaments, it is necessary to prepare protoplasts that can pass one at a time through the flow cell of the flow cytometer. We found that current flow cytometry methods for measuring GFP expression in plants were susceptible to debris resulting from protoplast preparations. We observed that when the plasma membrane of protoplasts is breached,

GFP diffuses out into the medium, and flow cytometric measurements of these non-viable protoplasts imply they do not express GFP. This debris can overestimate the proportion of non-expressing cells in the population. In order to correct this problem, we used an approach called a dye exclusion test. Because propidium iodide enters protoplasts to stain nuclei only when the plasma membrane is breached, debris that stains with this dye can be removed from our analysis. Using this approach we were able to quantitate GFP expression in individual cells without complications from debris.

We used flow cytometry to measure Green Fluorescent Protein (GFP) expression in individual tobacco NT-1 cells from lines transformed by *Agrobacterium*. We find that in this system the Rb7 MAR increases GFP expression 2-4 fold. This increase is caused by both an increase in the percentage of expressing cells and an increase in the magnitude of expression in expressing cells. Cell lines transformed with MAR-containing vectors averaged 27-39% more cells expressing GFP and these cells expressed GFP at 2-3 fold higher levels than cells transformed with control constructs. We also show that flow cytometry measurements on cells from isogenic lines are consistent with those from a population of cell lines obtained by liquid culture of entire *Agrobacterium* co-cultivation plates. By obviating the need to establish isogenic lines, this use of flow cytometry could greatly simplify the evaluation of MARs or other sequence elements that affect transgene expression. Our results indicate that the Rb7 MAR increases the frequency of transgene expression, presumably by reducing gene silencing, while also increasing the levels of expression in expressing cells, perhaps through an enhancer-like activity.

The Use of Flow Cytometry to Investigate the Effects of Matrix Attachment Regions on Transgene Expression in Plant Cells

by

Christopher Jay Halweg

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy

GENETICS

Raleigh

2004

APPROVED BY:

STEVEN SPIKER	WILLIAM THOMPSON
Co-chair of Advisory Committee	Co-chair of Advisory Committee
DOMINIQUE ROBERTSON	ARTHUR WEISSINGER

BIOGRAPHY

Christopher Jay Halweg was born on December 14, 1975 in Dubuque, Iowa and adopted five days later. He grew up both in and around Decorah, Iowa and actually lived on a farm for the first four years of his life, like all good Iowans should. He graduated from Decorah Senior High School in 1994.

With the intensions of going to school for only two years and finding a job, Chris enrolled into the Biotechnology program at Ellsworth Community College in Iowa Falls, Iowa. While attending Ellsworth, he was offered a USDA-ARS internship at Iowa State University for the summer of 1995. He was responsible for analyzing the fatty acid content of oil in maize germplasm as part of an international program called GEM (Germplasm Enhancement of Maize). In 1996, he graduated from Ellsworth Community College and transferred to Iowa State University in Ames, Iowa that fall.

While at Iowa State, Chris continued his internship with the USDA for the duration of his stay in Ames. In addition to his first real experience working in a laboratory setting, he also worked in the field during pollination season. Of course field work is a prerequisite to call yourself a true Iowan. At the same time he also worked in a molecular biology lab isolating genomic clones of Chloroplastic Acetyl-Coenzyme A Carboxylase in *Arabidopsis thaliana*. In 1998, he graduated from Iowa State University with a Bachelor's of Science degree in Genetics with a minor in Statistics.

Chris was admitted to the Genetics Department at North Carolina State University in 1998, as a candidate for the degree of Doctor of Philosophy. He has undertaken his research under the guidance of Dr. Steven Spiker and Dr. William F. Thompson.

ACKNOWLEDGEMENTS

First off, I would like to thank my family for their continuing words of encouragement throughout this process. I would also have to say that I would not be here if it were not for wonderful experiences I had while at Iowa State University. Specifically, I would like to thank Susan Duvick and Mack Shen for their guidance.

While at North Carolina State University, many people have contributed to my success. I would like to acknowledge my PhD advisor Steven Spiker for his support and untiring efforts as a true teacher. I would also like to acknowledge Bill Thompson for thoughtful discussions that were truly the impetus for much of this work. I wish to thank all other members of my committee past and present: Jon Swaffield, Dominique Robertson, and Arthur Weissinger. I also acknowledge the technical assistance I have received from Rita Abranches, George Allen, Anton Calloway, Tim Sit, and Bekir Ülker. Without the resources and guidance from Janet Dow at the NCSU Flow Cytometry Facility at the College of Veterinary Medicine, none of this work would be possible. I would also like thank my fellow graduate students: Michael Buck, Kirk Francis, Rick Hall, Jennifer Ingram, Jason Lowery, Luke Mankin, Randy Shultz, and Kathy Turner for their assistance and support. Additionally, I'm appreciative of the humble leadership of Stephanie Curtis and her open door policy. I'm grateful for the financial support I received from Tri-Agency, GAANN, and the Genetics Department. I'm also grateful for the many one-sided conversations I have had with NC State's Men's Basketball Program as we have both shared many high and low points over the years. Finally, I am truly indebted to Gretchen Feulner for her constant support and kind words of encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABA	abscisic acid
BUR	base unpairing region
CaMV	cauliflower mosaic virus
CTAB	hexadecyltrimethylammonium bromide
DNA	deoxyribonucleic acid
DSB	double strand break
FACS	fluorescence-activated cell sorter
FL1	fluorescence detector 1
FSC	forward-angle scatter
GFP	green fluorescent protein
GUS	β-glucuronidase
HMG	high mobility group
LB	left border
LIS	lithium diiodosalicylate
LUC	luciferase
MAR(s)	matrix attachment region(s)
MES	2-(N-Morpholino)ethanesulfonic acid sodium salt
mRNA	messenger ribonucleic acid
MS	Murashige and Skoog
NPTII	neomycin phosophotransferase
NT-1	Nicotiana tabacum
PCR	polymerase chain reaction
PI	propidium iodide
PMT	photomultiplier tube
PTGS	post-transcriptional gene silencing
RB	right border
rDNA	ribosomal deoxyribonucleic acid
RIGS	repeat induced gene silencing
RLU	relative light unit
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SAR(s)	scaffold sttachment region(s)
SDS	sodium dodecyl sulfate
siRNA(s)	small interfering ribonucleic acid(s)
smRS-GFP	soluble modified red-shifted green fluorescent protein
SSC	side-angle scatter
SSC	saline sodium citrate
TAE	tris-acetate ethylenediaminetetraacetic acid
T-DNA	transferred-deoxyribonucleic acid
TGS	transcriptional gene silencing
tRNA	transfer ribonucleic acid

CHAPTER 1

Introduction

The research presented in this dissertation describes an investigation of a DNA sequence called a Matrix Attachment Region (MAR) that may be involved in organizing chromatin structure and facilitating transgene expression in higher plants. The focus of this research was two fold. One aspect (the subject of chapter 3) was to characterize in greater detail the effect of matrix attachment regions on the variegation of transgene expression in plant cells. Unfortunately existing techniques used to study transgene expression variegation are often limited to qualitative data or have limited resolution to quantitate variegation. The second aspect of this research (the subject of chapter 2) was to modify and develop new tools to quantitate transgene expression in individual plant cells. Through the use of flow cytometry a more precise characterization of the effects of MARs on transgene expression has been determined setting the stage to further elucidate the mechanisms by which MARs may mediate these effects.

Variegated Expression

The use of reporter genes in transgenic studies in plants has given the plant biology community tools to monitor and evaluate a wide variety of cellular processes. Of particular interest in this work was to assess the effect of DNA sequences that may alter reporter gene expression. Simple quantitation of reporter gene expression is complicated by the observation that reporter gene expression can often be variegated. In other words,

some cells express the transgene while other cells may not. Additionally, variegation can occur where levels of reporter gene expression differ between cells. Most approaches to measuring transgene expression in plants fall into two major categories: histochemical approaches and extractive approaches. Histochemical approaches are not quantitative, but they can reveal cell-to-cell epigenetic variation in reporter gene expression. Approaches in which the product of the reporter gene is extracted and then measured are quantitative but they obscure epigenetic cell-to-cell differences. To gain maximal benefit from reporter gene studies, a method that could combine these approaches would be desirable. Such a method would allow quantitative measurements of cell-to-cell epigenetic variation. Chapter two in this dissertation demonstrates how flow cytometry can be used as a tool to quantitatively assess reporter gene expression variegation.

Why might the quantitative assessment of reporter gene variegation be important in understanding how regulatory elements like MARs affect gene expression? The observed increases in reporter gene expression in a population of cells could arise by two scenarios, which are not mutually exclusive. First, population increases in expression could occur by increasing the level (magnitude) of expression uniformly in all cells. Second, population increases in expression could occur by increasing the percentage of expressing cells in a population (likelihood) without increasing the levels of expression in any subgroup of the population. Finally, any combination is possible whereby both the percentage of expressing cells and the levels of expression in expressing cells may change. These scenarios cannot be distinguished using conventional approaches to measuring reporter gene expression unless expression is measured in single cells. As our

knowledge of regulatory elements grows, it appears that the dynamic organization of DNA and chromatin structure may play an important role. Regulatory elements that take part in altering chromatin structure may in turn alter gene expression giving rise to a new epigenetic state. Variegation in gene expression may be observed in cases where more than one epigenetic state is present. In order to begin to understand the mechanisms by which regulatory elements may affect variegation, we must first be able to quantitate this variegation.

Examples of regulatory elements that alter cell-to-cell differences in gene expression come mostly from the animal literature, where methods of flow cytometry analysis are well established. Theses examples sparked interest in whether regulatory elements in plants may also alter variegation. Studies combining flow cytometry analysis and regulatory elements such as transcriptional enhancers, which are orientation independent DNA sequences that can increase transcription at a distance from a promoter, have elegantly illustrated the possible effects regulatory elements can have on variegation. Ronai et al. (1999) were able to delete the Eμ core enhancer and/or flanking matrix attachment regions and switch region (Sμ) from the major intron of the immunoglobulin heavy locus in mouse cell lines. By monitoring gene expression with flow cytometry, the authors found that deletion of the Eμ core enhancer alone reduced population levels of expression. This reduction was not caused by reducing the levels of expression in all cells but by reducing the percentage of expressing cells over time.

Evidence that enhancers may actually alter variegation by preventing genes from being relocalized to areas of condensed chromatin was provided by Francastel et al.

(1999). The authors used site specific recombination to test various functional elements of the 5'HS2 globin enhancer at various sites in the genome of a human cell line. In general, cell lines not containing the complete 5'HS2 enhancer had a higher percentage of non-expressing cells. Furthermore transgenes in these non-expressing cells were heavily methylated and had fewer DNaseI hypersensitive sites, hallmarks of repressive chromatin. Further microscopic analysis of interphase nuclei confirmed that transgenic loci in non-expressing cells relocalized, in three-dimensional space, to centromeric heterochromatin, and the presence of the functional enhancer prevented this relocalization. The mechanisms by which enhancers increase transcription have been reevaluated with the advent of flow cytometry data into a new model in which enhancers act in a stochastic fashion to increase the probability that a regulated gene will be transcribed (Fiering et al., 2000).

These examples of the effects of enhancers on variegation of transgene expression remind us that we are not simply studying effects on a single cell, but rather a population of cells in which there may be epigenetic differences between cells. Transgene expression in this population may also be dynamic. The use of bulk biochemical assays to measure transgene expression obscures variegation and changes in variegation that may occur. In the discussion that follows on the effect of MARs on transgene expression in plants, it may be useful to consider how the results reported could be interpreted in the context of a population structure and how models for the function of MARs could be evaluated in this context.

Nuclear Architecture and its Role in Regulation of Gene Expression

One of the most fundamental and amazing feats of the eukaryotic nucleus is the means by which nearly 2 meters of DNA, in the case of humans and many crop plants, is organized and packaged to fit in a 10-25 micrometer in diameter space. Pikaard (1998) compared DNA packaging to taking several kilometers of very fine thread and fitting it into a sphere the size of a golf ball. Not only does the DNA need to be compacted as much as 50,000-fold, but the DNA must still be accessible to accommodate necessary functions of the cell such as transcription and replication. In order to accomplish this seemingly overwhelming task, several levels of compaction are employed. The first level of compaction was described early on as beads-on-a-string where 11 nm beads or nucleosomes were interspersed along the length of 2 nm wide DNA (Kornberg, 1974; Olins and Olins, 1974). Each nucleosome was composed of a histone octamer core that allowed ~146 bp of DNA to wrap around compacting DNA 6-fold. This DNA/histone octamer interaction is the primary functional unit of chromatin. These beads-on-a-string are further compacted into a 30 nm fiber by the linking histone H1. Although the structure of the 30 nm fiber has never been elucidated, a prominent model is the solenoid model, in which the nucleosome fiber coils about itself to form a solenoid with approximately 6 nucleosomes per turn (Finch and Klug, 1976; Thoma et al., 1979). At this stage DNA is compacted 40-fold.

An additional higher order level of compaction in which closely spaced loops of DNA radiating from a proteinaceous scaffold was observed by Paulson and Laemmli (1977). Following removal of histones from mitotic chromosomes, electron micrographs

showed DNA loops of 30 – 90 kb spilling out away from a scaffold shaped like the original mitotic chromosome. DNA at the base of these loops appears to be tethered to the scaffold and these sequences are called Scaffold Attachment Regions (SARs). If histones were not removed, loops consisting of 180-300 nucleosomes in 30 nm fibers were observed resulting in ~700 fold packaging of DNA (Marsden and Laemmli, 1979).

Interphase nuclei are also thought to contain a proteinaceous network called the nuclear matrix as originally identified by Berezney and Coffey (1974). High-salt extractions and DNase I were originally used to isolate these nuclear matrices, but under similar conditions apparent artifacts have been observed whereby denatured nuclear proteins formed filament like structures similar to the nuclear matrix (Tan et al., 2000). In an attempt to minimize such artifacts, low concentrations of chaotropic agents such as LIS (lithium diiodosalicylate) have been used to remove chromosomal proteins in nuclear matrix preparations (Mirkovitch et al., 1984; Hall et al., 1991). Alternatively, nuclei encapsulated in agarose can be depleted of chromatin without high-salt concentrations by electrophoresis revealing a nucleoskeleton resembling the nuclear matrix (Jackson and Cook, 1985). Nevertheless, controversy still exists whether or not the nuclear matrix is a true *in vivo* structure or simply an aggregate of proteins resulting from isolation procedures (Pederson, 1998; Martelli et al., 2002).

During the process of isolating nuclear matrices, removal of histones releases loops of DNA that radiate away from their attachment points to give the appearance of a halo (Gerdes et al., 1994). Cleavage and removal of these loops of DNA with either DNase I or restriction enzymes leaves a fraction of DNA that remains bound to the

nuclear matrix. These DNA sequences are called Matrix Attachment Regions (MARs) and are generally considered the same as Scaffold Attachment Regions (SARs) mentioned previously (Bode and Maass, 1988). MARs are operationally defined by their ability to bind isolated nuclear matrices *in vitro* and have been isolated from a variety of organisms, including fungi (Amati and Gasser, 1988), animals (Mirkovitch et al., 1984), and plants (Hall et al., 1991). MARs are generally AT rich (greater than 65% AT), which is thought to facilitate DNA unwinding at base unpairing regions (BURs) that may be involved in matrix binding (Bode et al., 1996). A database of MAR sequences and nuclear matrix proteins involved in DNA attachment has been assembled for public use called "S/MARt DB" (Liebich et al., 2002). Although MARs may contain several different sequence elements, it has been a challenge to use sequence information alone to identify DNA fragments with the capacity to bind to the nuclear matrix. Relying on the cumulative effect of sequence elements such as AT tracts, *Drosophila* topoisomerase II sites, DNA bending properties and others, algorithms have been produced to identify potential MAR sequences (Boulikas, 1995; Benham et al., 1997; Singh et al., 1997; Glazko et al., 2001). It is not surprising that these attempts have met with limited success, as not all of these MAR-related motifs correlate well with the ability to bind nuclear matrices *in vitro* (Michalowski et al., 1999).

MARs are thought to play a critical role in organizing domains of chromatin and may be involved in demarcating borders between domains of active and inactive chromatin structures. A mammalian nucleus contains 25 million nucleosomes that are organized into approximately 60,000 loops with an average size of 70 kb (Bode et al.,

2000). Small loop domains tend to be composed of active genes while large loop domains contain inactive regions (Gasser and Laemmli, 1987). So far, estimates of loop domain sizes are smaller in plants. For example, Avramova et al. (1995) described nine loops with an average size of 30 kb along a 280 kb region of a maize chromosome. In a different approach Michalowski et al. (1999) estimated loop domain sizes to be 17 kb in tobacco by dividing the genome size of tobacco by their estimate of the number of MAR binding sites in tobacco nuclear matrices. These observations are based on *in vitro* experiments and it is yet to be determined if these observations reflect the *in vivo* situation.

In an effort to demonstrate a link between both replication and transcriptional activity with loop domain structures, Gerdes et al. (1994) used in situ hybridization, pulse labeling and fluorescence microscopy to study nuclear halos. The authors found that newly synthesized DNA localized close to the nuclear matrix at the base of halos. Additionally, they discovered probes for actively expressed genes also localized close to the matrix, while probes for inactive genes produced long strings of signal extending out into the halo. Following extraction of nuclear matrices, Wei et al. (1999) observed newly synthesized DNA and RNA at non-overlapping sites bound throughout isolated matrices. These observations suggest the nuclear matrix and sequences that tether DNA to the nuclear matrix may play an important role in replication of DNA and transcription of genes.

Similarly, the positions of transcribed transgenes within nuclear halos isolated from transgenic tobacco plants transformed with or without the Rb7 MAR flanking a

GUS reporter gene were studied by Rita Abranches as reported by Allen et al. (2000). The transformed plants were previously characterized by Ülker et al. (1999) and were demonstrated to have increased levels of transgene expression when they were transformed with Rb7 MAR-containing constructs. MAR-containing transgenic loci remain close to the nuclear matrix in halo preparations while non-MAR-containing loci tend to be found extending out into the halo. These observations in combination with those of Gerdes et al. (1994) suggest a link between genes that are actively transcribed and their close association with the nuclear matrix.

Beyond the formation of loop domains, a higher level of DNA organization has been observed where individual chromosomes are compartmentalized into discrete territories within the nucleus (Cremer and Cremer, 2001). Croft et al. (1999) compared the localization patterns of human chromosomes 18 and 19. The authors discovered that the gene rich chromosome 19 is preferentially located more internally in the interphase nucleus and is more closely associated with the internal nuclear matrix than chromosome 18, which is less gene dense and is localized to nuclear periphery. Chromosome 19 (unlike chromosome 18) is also replicated very early in the cell cycle. Not only are these territories maintained in human cells, but their relative positions are also conserved during primate evolution suggesting a functionally relevant higher-order chromatin arrangement (Tanabe et al., 2002).

MAR Effects on Transgene Expression

In recent years, gene transfer into a wide variety of plant species has become almost common practice, but lack of predictable and reliable transgene expression has plagued efforts to fully exploit gene transfer technology for plant improvement and basic research. Initially, major differences in transgene expression between independent primary transformants were attributed to transgene silencing caused by integration of transgenes into or near heterochromatin regions, a phenomenon similar to "position effects" that have been well characterized in *Drosophila* (reviewed in Weiler and Wakimoto, 1995). Increases in transgene expression could be explained by integration near endogenous enhancers or regulatory elements. This idea is supported by the high frequency of enhancer trapping in plants (Beilmann et al., 1992; Sundaresan et al., 1995; Campisi et al., 1999). Simplistic extrapolation from these ideas leads to the hypothesis that, if transgene constructs were designed to create their own independent loop domains, transgene expression may be buffered from the effects of endogenous neighboring chromatin and/or regulatory sequences.

The inclusion of MARs in transgene constructs was first tested in chicken cells with the "A-element" from the chicken lysozyme locus (Stief et al., 1989). The A-element conferred a 10-fold increase in transgene expression from stably integrated transgenes, but no effect in transient expression experiments. In the presence of the A-element, transgene expression was copy-number dependent, i.e., more copies of the transgene gave proportionally higher levels of transgene expression. The authors suggested that the A-element conferred resistance to chromosomal position effects.

Following these results, the question remained whether the A-element would have the same effect in cells from other species or in whole organisms? These observations were subsequently corroborated in both rat fibroblasts and in transgenic mice (Bonifer et al., 1990; Phi-Van et al., 1990). Following this work, dissection of the 3 kb A-element revealed that various *cis* elements are required for copy number dependent and elevated levels of expression (Bonifer et al., 1994; Huber et al., 1994; Phi-Van and Stratling, 1996). Interestingly, sub-fragments that retained the ability to bind to the nuclear matrix were insufficient to confer copy number dependent expression (Phi-Van and Stratling, 1996). This result suggests that factors other than simply the ability to bind the nuclear matrix may have been involved in augmenting levels of gene expression. Clearly caution should be used when interpreting results from MAR experiments, since large fragments of DNA are often tested. These large fragments may contain internal sequence elements that are not associated with binding to the nuclear matrix but may be involved in altering transgene expression.

Stemming from optimism from early MAR experiments in animal systems reporting copy number dependent and position effect independent increases in transgene expression, several labs tested whether similar effects would be observed in plants, reviewed by Allen et al. (2000). An update of the current data available since 2000 on the effect of MARs on the levels and variation in transgene expression in plants is summarized in Table 1-1. Breyne et al. (1992) first reported the effects of MARs on transgene expression in plants. The authors isolated a MAR called "P1-SAR" from a soybean clone harboring a seed-specific lectin gene by testing binding to tobacco nuclear

matrices. In transgenic tobacco experiments with the soybean P1-SAR and the human β-globin MAR (Jarman and Higgs, 1988), no substantial increases in levels of transgene expression were observed. In fact a small decrease in expression was observed when constructs contained the P1-SAR, while constructs with the β-globin MAR had marginal increases in transgene expression. The P1-SAR reduced the variance in transgene expression by 60%. It should be noted that the authors did not include the few transgenic lines that had undetectable levels of β-glucuronidase (GUS) activity in their analysis. Additionally, the distribution of GUS activity for control lines was skewed with a higher percentage of low expressing lines. These results contrasted with that of early reports in animal systems in which elevated levels of transgene expression were observed (Stief et al., 1989; Bonifer et al., 1990; Phi-Van et al., 1990). Later experiments using the same P1-SAR in barley reported a 3.4 fold increase in transgene expression and a 4.6 fold reduction in the variance (Petersen et al., 2002).

The work by Breyne et al. (1992) was just the beginning of several reports of using MARs to flank transgenes in plants with varying effects on transgene expression levels and variation. Subsequent work in many cases has demonstrated an increase in transgene expression and/or a reduction in variability (Allen et al., 1993, 1996; Mlynarova et al., 1994, 1995; Han et al., 1997; Liu and Tabe, 1998; Ülker et al., 1999; Vain et al., 1999, 2002; Cheng et al., 2001; Mendu et al., 2001; Fukuda and Nishikawa, 2003; Mankin et al., 2003). The observations of increases in expression and/or reductions in variability have not, however, been consistent. Some reports indicate that MARs have no effect on transgene expression (van Leeuwen et al., 2001; Annadana et

al., 2002; De Bolle et al., 2003; Sidorenko et al., 2003) , and in at least two cases, a decrease in transgene expression has been reported (Breyne et al., 1992; Holmes-Davis and Comai, 2002).

The Rb7 MAR

The 1.2 kb Rb7 MAR is the MAR used in the research described in this dissertation. The Rb7 MAR was isolated from a region 3' to the root-specific tobacco gene *Rb7* (Conkling et al., 1990) by scanning across a 23 kb region surrounding the *Rb7* gene for fragments that bound to the nuclear matrix (Hall et al., 1991). The Rb7 MAR is AT-rich (73.2%), binds with high affinity for the nuclear matrix (80%), and has several DNA sequence motifs that have been found in other MARs (Michalowski et al., 1999). The number of binding sites for the Rb7 MAR in a tobacco nucleus has been estimated at 400,000 (Michalowski et al., 1999). Further dissection of the Rb7 MAR has revealed that a 750 bp region retains the majority of the binding affinity for nuclear matrices (Hall, unpublished). To date, no transgene expression experiments have been conducted with the 750 bp sub-fragment.

Along with having a high affinity for the nuclear matrix, the Rb7 MAR was shown to have dramatic effects on levels of transgene expression in tobacco cells in culture. Cell lines transformed with constructs containing the Rb7 MAR expressed the GUS reporter gene at a level 58-fold higher than controls, but the MAR had no effect on expression variability (Allen et al., 1996). This fold difference was much greater than previous reports of MAR effects in transgene experiments (reviewed by Allen et al.,

2000). Since then, the Rb7 MAR has been used in several different plant transformation systems including tobacco plants (Ülker et al., 1999), rice (Vain et al., 1999), cacao (Maximova et al., 2003), poplar (Han et al., 1997), white pine (Levee et al., 1999), and sorghum (Able et al., 1999).

Besides effects on reporter gene expression, the Rb7 MAR may also be able to increase transformation efficiencies possibly through increases in selectable marker expression. Han et al. (1997) observed such an effect with the Rb7 MAR, but others did not (Cheng et al., 2001; Maximova et al., 2003). Additionally, other MARs have been reported to increase transformation efficiencies (Mlynarova et al., 1995; Shimizu et al., 2001; Petersen et al., 2002). Similar to conflicting reports of increases in transgene expression summarized in Table 1-1 and by Allen et al. (2000), the effects of MARs on transformation efficiencies may depend on several factors such as transformation method, species recalcitrance, selection pressure, etc.

Gene Silencing and MARs

Low and variable transgene expression have plagued efforts in plant biotechnology. Early experiments with transgenic plants showed that they sometimes contained the complete transgene, but no protein was produced (Horsch et al., 1985). Similarly, some attempts to over produce endogenous proteins with transgenes actually led to reduced expression (Napoli et al., 1990). In the last few years there have been rapid advances in our understanding of the mechanisms of gene silencing that may have been involved in reduced and variable transgene expression (Wolffe and Matzke, 1999).

The phenotype of gene silencing is characterized by reduced levels of mRNA encoded by suppressed gene(s), and individual cases fall into two major mechanistic classes: 1) those in which mRNA level is regulated transcriptionally and 2) those in which it is regulated post-transcriptionally (Anandalakshmi et al., 1998). Although we are still far from understanding variability in the silencing of transgenes, great progress has been made in understanding transcriptional gene silencing (TGS), post transcriptional gene silencing (PTGS) and the interplay between the two (Carrington and Ambros, 2003; Grewal and Moazed, 2003; Matzke and Matzke, 2003; Schramke and Allshire, 2003). The roles of gene silencing in plants appear to be both in host defense and in developmental regulation. Host defense mechanisms may have evolved to protect plant genomes from foreign invasion of nucleic acids such as those contained in transposable elements, viroids, RNA and DNA viruses, and bacteria (Matzke et al., 2000). Plants have also used these same gene silencing mechanisms to regulate gene expression during development (Carrington and Ambros, 2003).

Tandem arrays of transgenes can be silenced by a mechanism whereby transgene repeats are able to form repressive heterochromatin, reviewed by Henikoff (1998). In *Drosophila*, as few as three copies of the mini-*white* eye pigment reporter gene demonstrated variegated phenotypes and heterochromatin formation and spreading (Dorer and Henikoff, 1994, 1997). Similar observations of repeat induced silencing associated with heterochromatin formation have been made in *Arabidopsis* (Assaad et al., 1993; Ye and Signer, 1996). Some evidence that MARs may help to resist gene silencing in small to moderately sized transgene arrays came from work using the yeast ARS1

MAR in tobacco suspension culture cells (Allen et al., 1993). DNA introduced by particle bombardment, the technique used in these experiments, often creates a single complex transgenic locus containing many copies of the transgene (Pawlowski and Somers, 1996). Allen et al. (1993) observed the ARS1 MAR increased overall transgene expression levels 12-fold. The increase was not copy number dependent above a threshold of transgene copies. The threshold in cell lines transformed with MAR-containing constructs appears to be higher than in control cell lines, suggesting that the ARS1 MAR may resist silencing associated with having additional copies of sequence homology at low to moderate copy number levels. Similar results were obtained with the Rb7 MAR where the highest expressing lines had low to moderate copy number levels (Allen et al., 1996). It remains unknown whether the observed higher levels of transgene expression at moderate copy numbers when transgenes are flanked by MARs are the result of MARs resisting potential pairing interactions associated with tandem arrays or some other mechanism of gene silencing.

Another indication that MARs may counteract gene silencing is evidenced by changes in the distribution of transgene expression in MAR containing lines compared to controls. MAR lines tend to have a reduced proportion of low expressing and/or non-expressing, silenced, lines. Brouwer et al. (2002) examined variation across a large number of transgenic lines. The authors noted fewer non-expressing lines in populations of transformants that contained either the yeast ARS1 MAR or the maize Adh1 5' MAR than among populations of control transformants. Based on these observations, they suggested that MARs increase the likelihood that an integrated gene will be expressed but

not the level of expression (Brouwer et al., 2002). Similarly, Mankin et al. (2003) observed a MAR induced reduction in the frequency of low expressing tobacco cell lines with the Rb7 MAR, but they also noted that when these low expressing cell lines were removed from the analysis, MAR lines still showed an increase in GUS expression. The data from Mankin et al. (2003) suggest there may be two mechanisms by which MARs increase overall levels of transgene expression. One mechanism may involve inhibition of gene silencing, but another may involve enhancement of transgene expression in expressing lines. Others have reported similar observations in which transgenes flanked by MARs reduce the proportion of non-expressing lines (Vain et al., 1999, 2002; Petersen et al., 2002).

MARs may also counteract gene silencing in progeny of transformed plants. Ülker et al. (1999) showed that tobacco plants transformed with the Rb7 MAR flanking the GUS reporter gene were more likely than control plants to maintain readily detectable levels of GUS expression in subsequent generations. Control plants frequently gave rise to progeny that were silenced. Thus, MARs may help stabilize transgene expression through meiosis. Similar results were observed in rice where both the tobacco Rb7 MAR and the yeast ARS1 MAR significantly improved the stability of transgene expression in both the T_0 and the T_1 generations (Vain et al., 1999, 2002). The authors noted that MARs did not eliminate gene silencing, but rather increased the frequency of stably expressing lines. In contrast, neither the P-MAR nor the *Adh1*-MAR from maize prevented gene silencing of the GUS transgene in progeny of transgenic maize plants (Sidorenko et al., 2003).

As mentioned previously, gene silencing is characterized as transcriptional gene silencing or post-transcriptional gene silencing. From the work of Ülker et al. (1999) in tobacco plants, it was unclear which form of gene silencing MARs may be counteracting. One characteristic of post-transcriptional gene silencing is that it can be overcome by crossing plants with silenced transgenes with plants expressing a viral suppressor of post-transcriptional gene silencing, such as P1/HC-Pro (Anandalakshmi et al., 1998). When plants carrying a silenced GUS transgene flanked by Rb7 MARs were crossed with P1/HC-Pro plants, expression was reactivated in progeny (Allen et al., 2000). In contrast, silenced transgenes not flanked by MARs were not reactivated when crossed with P1/HC-Pro plants (Ascenzi et al., 2003). These results suggest that the Rb7 MAR may counteract TGS and not PTGS, since crosses with suppressors of PTGS reactivated transgene expression.

In contrast to results that suggested the Rb7 MAR counteracts TGS, data from Mlynarova et al. (2003) suggest the A-element MAR counteracts PTGS. In order to make comparisons with and without MARs at the same transgenic locus, these authors used a Cre-*lox* site-specific recombination system to remove a 5' A-element (Mlynarova and Nap, 2003). Three transgenic tobacco lines containing the GUS transgene flanked by MARs exhibited stable GUS expression and were chosen for further analysis. Removal of the 5' A-element resulted in silenced GUS expression in 2 of the 3 lines. The reduction in GUS expression was especially noticeable when homozygous progeny lacking the A-element were compared to the same lines still containing the A-element. Furthermore, the relative amount of GUS transcript in nuclear run on assays was of the

same order of magnitude whether from active or silenced plants suggesting PTGS. Additionally, small interfering RNAs (siRNAs), which are thought to target degradation of homologous RNA sequences in the PTGS pathway, were also present in silenced lines. These results give the most direct evidence that the A-element may counteract PTGS.

Often during the transformation of plants, multiple copies integrate into a single locus. In these cases gene silencing interactions are in *cis*. In order to evaluate the effect of MARs on the inhibition of *trans*-silencing mediated by sequence homologies elsewhere in the genome, Vaucheret et al. (1998) made crosses between transgenic tobacco plants. Several transgenic lines containing transgenes flanked by various different MARs (Mlynarova et al., 1994; van der Geest et al., 1994; Ülker et al., 1999) were crossed with either the strong transcriptional *trans*-silencing line V271 (Vaucheret, 1993) or the strong post-transcriptional *trans*-silencing line 6b8 (Elmayan and Vaucheret, 1996). The V271 line carries a single extensively rearranged multicopy locus containing the 35S promoter and the nitrate reductase coding region. This locus is known to transcriptionally silence other 35S promoters. Similarly, the 6b8 line carries a single locus containing the double 35S promoter driving the GUS reporter gene, and this locus is known to silence other GUS reporter genes by a post-transcriptional mechanism. None of the MARs tested counteracted either transcriptional or post-transcriptional *trans*-silencing when crosses were made to either of the strong silencing lines V271 or 6b8 (Vaucheret et al., 1998).

While Vaucheret et al. (1998) had shown that the Rb7 MAR did not prevent strong *trans*-silencing, it remained to be determined if the Rb7 MAR could protect

against weaker *trans*-silencing loci. Ascenzi et al. (2003) characterized two transcriptional *trans*-silencing lines C190 and C40 that were generated previously (Ülker et al., 1999). These new silencer lines were weaker at reducing expression in *trans* than the previously characterized V271 line (Vaucheret, 1993) with line C40 being the weakest. When tobacco lines containing Rb7 MAR flanked transgenes or controls without MARs were crossed with the strong *trans*-silencer V271 line, transgene expression was lost as observed previously (Vaucheret et al., 1998). In contrast, four out of six lines containing Rb7 flanked transgenes showed reduced sensitivity to the weaker *trans*-silencing lines C190 and C40 compared to control crosses. While MARs do not completely eliminate *trans*-silencing, these results suggest MARs may reduce the sensitivity of a locus to less severe forms of trans-silencing.

There appears to be no straight forward mechanistic explanation for how MARs may affect gene silencing. Several lines of evidence do suggest that MARs counteract the loss of transgene expression demonstrated by either a reduction in transgene expression variability or an increase in overall transgene expression (Table 1-1 and Allen et al., 2000). In a few studies, a reduced proportion of low expressing and/or non-expressing, silenced, lines have been observed when MARs flank transgenes (Brouwer et al., 2002; Petersen et al., 2002; Mankin et al., 2003). Additionally, MARs increase the stability of expression of transgenes in subsequent generations (Ülker et al., 1999; Vain et al., 1999, 2002). MARs cannot overcome strong *trans*-silencing effects (Vaucheret et al., 1998), but reduce sensitivity to weak *trans*-silencing (Ascenzi et al., 2003). Whether these overall effects on gene silencing mediated by MARs are at the transcriptional or

post-transcriptional level are yet to be fully understood, but evidence thus far suggests the Rb7 MAR may counteract TGS (Allen et al., 2000; Ascenzi et al., 2003) and the A-element may counteract PTGS (Mlynarova et al., 2003). More studies will be required to understand how these contrasting observations relate to general aspects of MAR function in plants.

Modeling the MAR Effect

While there have been many experiments investigating MARs in plants, there is little evidence on how MARs may function. Early on, much of the focus was on determining what combination of MAR flanking sequences in various transgenic host species would yield high, less variable levels of transgene expression. There is no formula to predict to what extent a given MAR will augment levels of transgene expression as various reports have had conflicting conclusions. However, the preponderance of the data are consistent with the notion that MARs often increase transgene expression or reduce its variability, or both, depending on additional variables that remain to be understood.

MARs are defined by their capacity to bind nuclear matrices *in vitro*, but effects on transgene expression and variation are *in vivo*. It remains to be determined if MARs truly bind the nuclear matrix *in vivo* and if it is this binding that affects gene expression. Another confounding factor is that identified MAR regions are often large pieces of DNA that may have other sequence elements embedded in them (Bonifer et al., 1994; Phi-Van and Stratling, 1996). Since MARs cannot be identified based on sequence alone (Glazko

et al., 2001), it can be a laborious task (involving actual matrix-binding assays) to distinguish DNA sequences that facilitate matrix binding from those that have other effects on transgene expression. Models of MAR function have been discussed in detail (Allen et al., 2000). The models summarized here include several in which matrix binding *in vivo* is not considered essential.

Independent Loop Domain

MARs may affect transgene expression by creating independent loop domains that are topologically isolated from neighboring chromatin that has a transcriptionally inactive conformation. Unfortunately, this can be a very difficult model to test since matrix binding has yet to be demonstrated *in vivo*. Nevertheless, *in situ* hybridization has shown that active genes tend to localize close to the matrix, while inactive genes tend to be localized in the surrounding halo DNA (Gerdes et al., 1994). Similarly, work by Rita Abranches as reported by Allen et al. (2000) demonstrated transgenic loci containing the Rb7 MAR were also localized more closely to the nuclear matrix than were transgenes not associated with the Rb7 MAR. These transgenes flanked by MARs tended to have higher levels of expression than controls (Ülker et al., 1999).

If flanking a transgene with MARs actually creates an independent loop domain, then why should the domain have an active chromatin structure verses and inactive chromatin structure? One possible explanation is that the putative loop formed in many of the previous transgenic experiments would have been too small to support a 30 nm chromatin fiber structure. Given this possibility, Mendu et al. (2001) added 6.6 kb of

lambda phage spacer DNA between cloned Rb7 MARs to create a putative loop domain that they argued was large enough to form a 30 nm chromatin fiber structure. The authors observed no difference in transgene expression whether the putative loop domain was 3 kb verses 9.6 kb.

In contrast to Mendu et al. (2001), Cheng et al. (2001) observed when a Rb7 MAR bounded putative loop domain's size was increased, there was a reduction in transgene expression levels, and the variability in transgene expression increased. The authors compared a 4.6 kb putative loop with a 6.9 kb putative loop measuring GFP fluorescence and compared a 5.7 kb putative loop with an 8.0 kb putative loop measuring GUS activity. It's not clear whether the larger putative loops may support a 30 nm structure, even so inclusion of the additional lambda spacer DNA did lower transgene expression and increase variability in transgene expression suggesting either loop size or gene cassette position within the loop may have an effect. Transgenes positioned in the middle of a putative loop bounded by the A-element exhibited higher levels of expression than transgenes positioned near the 3' end of the putative loop (Mlynarova et al., 2002).

Chromatin Opening and Chromatin Remodeling

The chromatin opening model suggests that titrating away histone H1, which stabilizes 30 nm chromatin fibers, will lead to local opening of chromatin that will spread to adjacent regions (Kas et al., 1993). Histone H1 preferentially binds to AT-rich DNA and specifically associates with MARs nucleating H1 assembly into neighboring DNA (Izaurralde et al., 1989). By titrating away histone H1 with the drug distamycin,

characteristics of decondensed chromatin were observed (Kas et al., 1993), suggesting other biological molecules may be able to remodel chromatin at MAR sequences. The non-histone chromatin HMG I/Y proteins also preferentially bind AT-rich DNA and can displace histone H1 from MAR sequences (Zhao et al., 1993). In mammalian systems, HMG I/Y proteins are more abundant in rapidly dividing, undifferentiated cells (Johnson et al., 1990). Since Rb7 MAR effects on transgene expression were much greater in rapidly dividing cell lines (Allen et al., 1996) than in whole plants (Ülker et al., 1999), it was hypothesized that these differences may be attributed to different levels of HMG I/Y. Later analysis of tobacco cell cultures initiated from these same transgenic tobacco plants did not demonstrate that MAR effects were enhanced by initiation of rapidly dividing cell cultures (Ascenzi et al., 2001). Furthermore the authors observed that levels of HMG I/Y protein did not significantly differ between tobacco plants and tobacco culture cells.

Chromatin remodeling around MARs may facilitate an enhancer's ability to interact with promoters and boost expression. Insulators are DNA sequences that are defined by their capacity to block the effect of enhancers when placed between an enhancer and a regulated promoter. While testing for insulator function of the bean β-phaseolin 5' MAR in tobacco plants, van der Geest and Hall (1997) observed the β-phaseolin 5' MAR actually facilitated the 35S enhancer element activation of GUS expression, while other similarly sized DNA fragments reduced GUS expression to basal levels. The β-phaseolin 5' MAR did not function as an enhancer by itself in transient assays. The authors suggested the β-phaseolin 5' MAR may mediate its effects by altering chromatin structure. Also, the presence of the enhancer from the pea

plastocyanin gene (*PetE*), which binds to nuclear matrices as well, correlates with histone acetylation of downstream promoter and coding regions suggesting MARs may be targets for histone acetyltransferases and/or may be involved in maintaining an open chromatin structure (Chua et al., 2003). Similarly, MARs of the immunoglobulin μ gene are required to extend a domain of histone acetylation for long range interactions between an intragenic μ enhancer and a distal promoter (Forrester et al., 1994, 1999; Fernandez et al., 2001).

The identification of proteins that bind to MARs may play an important role in understanding how MARs affect gene expression as reviewed by Martens et al. (2002) and Bode et al. (2003). The matrix-localized, scaffold-attachment factor A (SAF-A) contains a 31 amino acid domain called a SAF-box that mediates binding to MARs (Kipp et al., 2000). SAF-A alone has weak affinity for MARs, but as more and more SAF-A proteins bind, SAF-A's affinity for MARs increases. SAF-A also binds a histone acetyltransferase (p300) resulting in acetylation of local histones (Martens et al., 2002). Actetylated histones are associated with an open chromatin state, suggesting that SAF-A may act as a bridge between MARs and chromatin remodeling machinery.

SATB1 (special AT-rich sequence binding 1), another MAR binding protein, can assemble either a transcriptionally active complex or a silencing complex depending on other host factors that are recruited during T-cell differentiation (Bode et al., 2003). A unique cage-like network of SATB1 surrounding areas of dense chromatin has been observed in nuclei of thymocytes (Cai et al., 2003). It's thought that this SATB1 network may serve as a landing platform for chromatin-remodeling factors at the bases of

chromatin loops and that these factors could regulate chromatin structure and gene expression over long distances (Yasui et al., 2002; Cai et al., 2003).

Influencing Integration

Technological advances over the past several years have enabled the transformation of several plant species without direct knowledge of the mechanisms of transgene integration. Both *Agrobacterium* and direct DNA delivery methods of transformation are thought to occur through illegitimate recombination at microhomologies of 1-8 bp, reviewed by Somers and Makarevitch (2004). Furthermore, transgene DNA may integrate at double stranded breaks (DSB) through DSB repair (Gorbunova and Levy, 1997; Pawlowski and Somers, 1998; Salomon and Puchta, 1998; Vergunst and Hooykaas, 1999). Characteristics of MARs such as AT-rich sequences or regions which contain base unwinding elements are also more prone to breaks. Indeed, MARs or putative MARs have been identified in border junctions of transgenic loci in *Arabidopsis* (Sawasaki et al., 1998), oat (Makarevitch et al., 2003), rice (Takano et al., 1997), and tobacco (Shimizu et al., 2001). It's unclear whether flanking a transgene with MARs could help facilitate integration or possiblly direct transgenes to more active regions of the genome. The latter is difficult to test since there may be selection bias for active regions of the genome, and plants with transgenes integrating into inactive regions of the genome may not be recovered. Recent observations in *Arabidopsis* suggest that T-DNA integration is random when selection is not a prerequisite for identification of

transgenic plants (Francis, 2004). Similar approaches may be necessary to elucidate the manner in which MARs may influence the integration of transgenes.

MARs as Terminators

MARs may augment levels of transgene expression by acting as additional transcription terminator sequences that may reduce the negative effects of read-through into or from neighboring transcriptional units. Transcriptional interference can be caused by invading transcription from either upstream or downstream promoters to reduce gene expression (Ingelbrecht et al., 1989, 1991; Thompson and Myatt, 1997; Greger and Proudfoot, 1998; Padidam and Cao, 2001; Valerius et al., 2002). Possible mechanisms of transcriptional interference include antisense RNA production, RNA polymerase collision, and DNA topological constraints (Valerius et al., 2002). Villemure et al. (2001) demonstrated that the A-element MAR was able to protect against transcriptional interference in mouse cells, but only when the MAR was positioned between a head-to-tail arrangement of transgenes. Plant transformation often involves the generation of multicopy loci frequently containing strong constitutive promoters. The formation of inverted repeats is likely and could produce antisense RNAs that are known to be strong elicitors of gene silencing (Muskens et al., 2000). Even read-through transcription of promoter sequences have been shown to silence homologous promoters (Mette et al., 1999). MARs may play a role in helping to terminate transcripts before invading neighboring transcriptional units. Such termination may be due to the MARs being AT-rich or physical blocking of RNA polymerases by interactions with the nuclear matrix.

Although MARs may act as terminators, it's clear not all sequences with the capacity to bind the nuclear matrix *in vitro* act as terminators. MARs have been discovered within introns of genes (Kas and Chasin, 1987; Cockerill, 1990; Romig et al., 1994) suggesting that there are at least some circumstances when MARs would not function as terminators. MARs often have extensive base unwinding regions that may relieve super helical strain (Bode et al., 1992, 1996; Kay and Bode, 1994). An alternative hypothesis of how MARs may be able to reduce the effects of transcriptional interference is by acting as a sink to absorb positive supercoils generated in front of an advancing polymerase transcriptional complex. The idea that MARs may relieve super helical strain has been suggested as a function for intronic MARs (Bode et al., 1992).

Carpet Tack Model

Repeat Induced Gene Silencing (RIGS) arises from tandem arrays of transgenes that may interact in *cis* to form condensed, repressive heterochromatin (Henikoff, 1998). Since the transformation of plants can often generate multicopy transgenic loci, there may be several opportunities for DNA-DNA pairing interactions among homologous repeat sequences. MARs may inhibit these interactions by 'tacking' transgene arrays to the nuclear matrix and prevent homology sensing mechanisms (Allen et al., 2000). Evidence for heterochromatin formation during RIGS has been well established in *Drosophila* (Dorer and Henikoff, 1994, 1997), mice (Garrick et al., 1998) and *Arabidopsis* (Assaad et al., 1993; Ye and Signer, 1996).

While transcriptional interference may also explain silencing of transgene repeats, evidence does exist in plants that homologous transgenes may actually pair and/or localize to the same nuclear position (Assaad and Signer, 1992; Abranches et al., 2000; Kohli et al., 2003). Of particular interest was the observation of homologous recombination within a transgene locus between two non-functional alleles of the *npt* gene to produce a functional *npt* gene and confer resistance to kanamycin in *Arabidopsis* (Assaad and Signer, 1992). Although this observation is not direct evidence of DNA-DNA pairing, it does suggest an interaction between regions of homology may exist. Furthermore, plants from these recombination events generated single inserts as well as tandem inserts. Nearly all plants that exhibited silenced phenotypes contained tandem inserts (Assaad et al., 1993). Plants with these tandem inserts demonstrated RIGS and heterochromatin formation of transgene arrays (Ye and Signer, 1996).

While MARs may prevent interactions among transgene repeats, MARs do not prevent gene silencing at high transgene copy numbers (Allen et al., 1993, 1996). One possible explanation for this phenomenon is that transgenes flanked by MARs may produce quantities of mRNA that increase with transgene copy number until a threshold is attained. When mRNA levels exceed the threshold, post-transcriptional silencing may be triggered, as proposed by Allen et al. (2000). Others have suggested that silencing associated with repetitive DNA structures may be attributed to high transgene doses rather than the repeat arrangements themselves (Lechtenberg et al., 2003). Whether MARs prevent transgene interactions or buffer transgene copy numbers up to a threshold value, clearly plants and other organisms have evolved mechanisms to escape RIGS as

evidenced by the fact that high copy number repeated genes such as histone, rRNA, and tRNA genes are not silenced.

Advances in Understanding the Influences of MARs on Expression Variegation

In the work presented in this dissertation, variegation in transgene expression was observed, as has been previously reported (Allen et al., 1993; Ülker et al., 1999). In order to study the possible mechanisms by which MARs may alter the expression of transgenes, we must first accurately quantify transgene expression and transgene expression variegation. Since MARs appear to buffer the effects of gene silencing and since gene silencing by its nature is epigenetic, it will be important to quantify these epigenetic effects. In chapter two of this dissertation I discuss and illustrate methods to quantify such epigenetic effects through the use flow cytometry to quantify GFP expression in individual plant cells.

In chapter three of this dissertation I report on the use of flow cytometry to quantify GFP expression and variegation effects attributable to flanking a GFP reporter gene with the Rb7 MAR. The Rb7 MAR appears to have a dual impact on transgene expression, acting both by increasing the probability or likelihood that a cell will express a transgene and by increasing the magnitude of expression in expressing cells. In other words, cell lines containing transgenes flanked by MARs frequently have more cells expressing the transgene, and these cells express at higher levels. These results suggest there may be two modes of action whereby population increases in transgene expression are observed. First, the propensity for transgene expression may be stochastic in nature with a given

probability of an expression compatible state. The Rb7 MAR may simply increase the probability of an expression compatible state. Whether this expression compatible state may be the result of counteracting gene silencing mechanisms and/or altering chromatin structure or some other mechanism is unknown. Second, once an expression compatible state is achieved, the Rb7 MAR mediates an increase in the levels of expression in the expressing population. Increases in transcription could occur, for example, if MARs bring transgenes close to matrix-associated transcriptional machinery, if they relieve torsional stress associated with advancing polymerases, if they facilitate long range interactions with endogenous enhancers, or if they promote remodeling of local chromatin structure. The capability to identify and quantitate subpopulations within transgenic lines, established in this work, will greatly facilitate future investigations of these and other potential mechanisms. Additionally, these flow cytometry techniques will be of use in studies of the modes of action of other regulatory elements that affect transcription.

Table 1-1. MAR Mediated Effects on Transgene Expression Reported Since 2000.

Reference	System	MAR Source	MAR Name	Promoter-Reporter	Transformation Method	Expression Level Increase	Effect on Expression Variability
Annadana et al., 2002	Chrysanthemum Plants	Chicken	A-element	35s2x-GUS	Agrobacterium	None	None
Annadana et al., 2002	Chrysanthemum Plants	Chicken	A-element	Lhca3-GUS	Agrobacterium	None	Slight Increase
Ascenzi et al., 2003	Tobacco Plants	Tobacco	Rb7	35s-GUS	Biolistic	Slight	Not Tested
Brouwer et al., 2002	Maize Cells	Yeast	ARS1	35s-LUC	Biolistic	5.8x	None
Brouwer et al., 2002	Maize Cells	Maize	Adh1 5'	35s-LUC	Biolistic	26.8x	None
Brouwer et al., 2002	Maize Cells	Maize	Mha1 5'	35s-LUC	Biolistic	1.5x	None
Brouwer et al., 2002	Maize Plants	Maize	Adh1 5'	Rsyn7-GUS	Biolistic	87.0x Decrease	Not Tested
Cheng et al., 2001	Rice Plants	Tobacco	Rb7	ACT1-GFP	Biolistic	18.4x; 3.3x[1]	Slight Reduction
Cheng et al., 2001	Rice Plants	Tobacco	Rb7	ACT1-GUS	Biolistic	650x; 376x[1]	Slight Reduction
De Bolle et al., 2003	Arabidopsis Plants	Chicken	A-element	35s-GUS	Agrobacterium	None	None
De Bolle et al., 2003	Arabidopsis Plants	Chicken	A-element	MAS-GUS	Agrobacterium	None	None
Fukuda and Nishikawa, 2003	Tobacco Cells	Tobacco	5' CHN50	35s-GUS	Agrobacterium	15x; 10x[2]; None[3]	None
Holmes-Davis and Comai, 2002	Arabidopsis Plants	Tomato	HSC80 5' & 3'	APT-Lc cDNA	Agrobacterium	None	Not Tested
Holmes-Davis and Comai, 2002	Arabidopsis Plants	Yeast	ARS1	APT-Lc cDNA	Agrobacterium	Slight Decrease	Not Tested
Mankin et al., 2003	Tobacco Cells	Tobacco	Rb7	35s-GUS	Biolistic	5.3x	5x Fewer Non-expressers
Mankin et al., 2003	Tobacco Cells	Tobacco	Rb7	AtAhas-GUS	Biolistic	None	5x Fewer Non-expressers
Mankin et al., 2003	Tobacco Cells	Tobacco	Rb7	PsFed1-GUS	Biolistic	None	None
Mankin et al., 2003	Tobacco Cells	Tobacco	Rb7	GmHspL-GUS	Biolistic	3.1x	None
Mankin et al., 2003	Tobacco Cells	Tobacco	Rb7	Nos-GUS	Biolistic	4.9x	5x Fewer Non-expressers
Mankin et al., 2003	Tobacco Cells	Tobacco	Rb7	OCS-GUS	Biolistic	15.3x	5x Fewer Non-expressers
Maximova et al., 2003	Cacao Plants	Tobacco	Rb7	E12omega-GFP	Agrobacterium	2.2x	Reduction
Mendu et al., 2001	Tobacco Cells	Tobacco	Rb7	35s-GUS	Biolistic	42x	Not Tested
Mlynarova et al., 2002	Tobacco Plants	Chicken	A-element	35s-LUC	Agrobacterium	None	1-15x Reduction
Mlynarova et al., 2002	Tobacco Plants	Chicken	A-element	Lhca3-GUS	Agrobacterium	None	1-18x Reduction
Mlynarova et al., 2003	Tobacco Plants	Chicken	A-element	35s2x-GUS	Agrobacterium	None	Not Tested
Nowak et al., 2001	Tobacco Plants	Synthetic	sMAR	35s-GUS	Agrobacterium	2.3x; 3.8x[1]	Slight Increase
Petersen et al., 2002	Barley Plants	Soybean	P-1MAR	35s-GUS	Biolistic	3.4x[4]	4.6x Reduction
Petersen et al., 2002	Barley Plants	Petunia	TBS-MAR	35s-GUS	Biolistic	None	None
Shimizu et al., 2001	Tobacco Cells	Tobacco	TJ1	35s-NPTII	Biolistic	Slight	Not Tested
Sidorenko et al., 2003	Maize Plants	Maize	P-MAR	(P1-rr, WP, Rsyn7)-GUS	Biolistic	None	None
Sidorenko et al., 2003	Maize Plants	Maize	Adh1-MAR 5'	(P1-rr, WP, Rsyn7)-GUS	Biolistic	None	None
Vain et al., 2002	Rice Plants	Tobacco	Rb7	35s-GUS	Biolistic	2.5x	None
Vain et al., 2002	Rice Plants	Yeast	ARS1	35s-GUS	Biolistic	3.1x	None
van der Geest et al., 2004	Rice Plants	Synthetic	SM11	GOS2-GUS	Biolistic	None; 1.9x[4]	None
van der Geest et al., 2004	Arabidopsis Plants	Synthetic	SM11	ACT2-GUS	Agrobacterium	3.7x; 1.6x[4]	None
van Leeuwen et al., 2001	Tobacco Plants	Chicken	A-element	35s2x-LUC	Agrobacterium	None; 2.0x Decrease[5]	None; Increase[5]
Zhang et al., 2002	Tobacco Plants	Tobacco	TM1	35s-GUS	Agrobacterium	1.5x	Not Tested
Zhang et al., 2002	Tobacco Plants	Tobacco	TM2	35s-GUS	Agrobacterium	5x	Not Tested
Zhang et al., 2002	Tobacco Plants	Tobacco	TM3	35s-GUS	Agrobacterium	1.4x	Not Tested
Zhang et al., 2002	Tobacco Plants	Arabidopsis	AM1	35s-GUS	Agrobacterium	1.3x	Not Tested
Zhang et al., 2002	Tobacco Plants	Arabidopsis	AM2	35s-GUS	Agrobacterium	None	Not Tested

[1]Additional spacer DNA was included in the transgene construct.

[2]A single MAR was positioned only 5' to the transgene but tested in both possible orientations.

[3]A single MAR was positioned only 3' to the transgene.

[4]A single MAR was positioned both 5' and 3' flanking the transgene as indirect repeats.

[5]The reporter gene was positioned near the 3' MAR.

References

Able, J.A., Rathhus, C., Carroll, B.J., and Godwin, I.D. (1999). Enhancing transgene expression levels in sorghum: Current status and future goals. (International Crops Research Institute for the Semi-Arid Tropics & Oxford Publilshers).

Abranches, R., Santos, A.P., Wegel, E., Williams, S., Castilho, A., Christou, P., Shaw, P., and Stoger, E. (2000). Widely separated multiple transgene integration sites in wheat chromosomes are brought together at interphase. Plant J **24,** 713-723.

Allen, G.C., Spiker, S., and Thompson, W.F. (2000). Use of matrix attachment regions (MARs) to minimize transgene silencing. Plant Mol Biol **43,** 361-376.

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

Allen, G.C., Hall, G., Jr., Michalowski, S., Newman, W., Spiker, S., Weissinger, A.K., and Thompson, W.F. (1996). High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. Plant Cell **8,** 899-913.

Amati, B.B., and Gasser, S.M. (1988). Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold. Cell **54,** 967-978.

Anandalakshmi, R., Pruss, G.J., Ge, X., Marathe, R., Mallory, A.C., Smith, T.H., and Vance, V.B. (1998). A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A **95,** 13079-13084.

Annadana, S., Mlynarova, L., Udayakumar, M., de Jong, J., and Nap, J.P. (2002). The potato Lhca3.St.1 promoter confers high and stable transgene expression in chrysanthemum, in contrast to CaMV-based promoters. Mol Breeding **8,** 335-344.

Ascenzi, R., Ingram, J.L., Massel, M., Thompson, W.F., Spiker, S., and Weissinger, A.K. (2001). The role of cell differentiation state and HMG-I/Y in the expression of transgenes flanked by matrix attachment regions. Transgenic Res **10,** 465-470.

Ascenzi, R., Ülker, B., Todd, J.J., Sowinski, D.A., Schimeneck, C.R., Allen, G.C., Weissinger, A.K., and Thompson, W.F. (2003). Analysis of trans-silencing interactions using transcriptional silencers of varying strength and targets with and without flanking nuclear matrix attachment regions. Transgenic Res **12,** 305-318.

Assaad, F.F., and Signer, E.R. (1992). Somatic and germinal recombination of a direct repeat in Arabidopsis. Genetics **132,** 553-566.

Assaad, F.F., Tucker, K.L., and Signer, E.R. (1993). Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. Plant Mol Biol **22,** 1067-1085.

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

Beilmann, A., Albrecht, K., Schultze, S., Wanner, G., and Pfitzner, U.M. (1992). Activation of a truncated PR-1 promoter by endogenous enhancers in transgenic plants. Plant Mol Biol **18,** 65-78.

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

Berezney, R., and Coffey, D.S. (1974). Identification of a nuclear protein matrix. Biochem Biophys Res Commun **60,** 1410-1417.

Bode, J., and Maass, K. (1988). Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. Biochemistry **27,** 4706-4711.

Bode, J., Stengert-Iber, M., Kay, V., Schlake, T., and Dietz-Pfeilstetter, A. (1996). Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. Crit Rev Eukaryot Gene Expr **6,** 115-138.

Bode, J., Goetze, S., Heng, H., Krawetz, S.A., and Benham, C. (2003). From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. Chromosome Res **11,** 435-445.

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

Bode, J., Benham, C., Ernst, E., Knopp, A., Marschalek, R., Strick, R., and Strissel, P. (2000). Fatal connections: when DNA ends meet on the nuclear matrix. J Cell Biochem Suppl **35,** 3-22.

Bonifer, C., Vidal, M., Grosveld, F., and Sippel, A.E. (1990). Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. EMBO J **9,** 2843-2848.

Bonifer, C., Yannoutsos, N., Kruger, G., Grosveld, F., and Sippel, A.E. (1994). Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. Nucleic Acids Res **22,** 4202-4210.

Boulikas, T. (1995). Chromatin domains and prediction of MAR sequences. Int Rev Cytol **162A,** 279-388.

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

Brouwer, C., Bruce, W., Maddock, S., Avramova, Z., and Bowen, B. (2002). Suppression of transgene silencing by matrix attachment regions in maize: a dual role for the maize 5' ADH1 matrix attachment region. Plant Cell **14,** 2251-2264.

Cai, S., Han, H.J., and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nat Genet **34,** 42-51.

Campisi, L., Yang, Y., Yi, Y., Heilig, E., Herman, B., Cassista, A.J., Allen, D.W., Xiang, H., and Jack, T. (1999). Generation of enhancer trap lines in Arabidopsis and characterization of expression patterns in the inflorescence. Plant J **17,** 699-707.

Carrington, J.C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. Science **301,** 336-338.

Cheng, Z.Q., Targolli, J., and Wu, R. (2001). Tobacco matrix attachment region sequence increased transgene expression levels in rice plants. Mol Breeding **7,** 317-327.

Chua, Y.L., Watson, L.A., and Gray, J.C. (2003). The transcriptional enhancer of the pea plastocyanin gene associates with the nuclear matrix and regulates gene expression through histone acetylation. Plant Cell **15,** 1468-1479.

Cockerill, P.N. (1990). Nuclear matrix attachment occurs in several regions of the Igh locus. Nucleic Acids Res **18,** 2643-2648.

Conkling, M.A., Cheng, C.L., Yamamoto, Y.T., and Goodman, H.M. (1990). Isolation of transcriptionally regulated root-specific genes from tobacco. Plant Physiol **93,** 1203-1211.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet **2,** 292-301.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol **145,** 1119-1131.

De Bolle, M.F.C., Butaye, K.M.J., Coucke, W.J.W., Goderis, I.J.W.M., Wouters, P.F.J., van Boxel, N., Broekaert, W.F., and Cammue, B.P.A. (2003). Analysis of the influence of promoter elements and a matrix attachment region on the inter-individual variation of transgene expression in populations of Arabidopsis thaliana. Plant Sci **165,** 169-179.

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

Dorer, D.R., and Henikoff, S. (1997). Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. Genetics **147,** 1181-1190.

Elmayan, T., and Vaucheret, H. (1996). Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally. Plant J **9,** 787-797.

Fernandez, L.A., Winkler, M., and Grosschedl, R. (2001). Matrix attachment region-dependent function of the immunoglobulin mu enhancer involves histone acetylation at a distance without changes in enhancer occupancy. Mol Cell Biol **21,** 196-208.

Fiering, S., Whitelaw, E., and Martin, D.I. (2000). To be or not to be active: the stochastic nature of enhancer action. Bioessays **22,** 381-387.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A **73,** 1897-1901.

Forrester, W.C., Fernandez, L.A., and Grosschedl, R. (1999). Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. Genes Dev **13,** 3003-3014.

Forrester, W.C., van Genderen, C., Jenuwein, T., and Grosschedl, R. (1994). Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. Science **265,** 1221-1225.

Francastel, C., Walters, M.C., Groudine, M., and Martin, D.I.K. (1999). A functional enhancer suppresses silencing of a transgene and prevents its localization close to centromeric heterochromatin. Cell **99,** 259-269.

Francis, K.E. (2004). The effects of T-DNA integration sites on transgene expression in Arabidopsis [dissertation]. In Genetics (Raleigh, NC: North Carolina State University).

Fukuda, Y., and Nishikawa, S. (2003). Matrix attachment regions enhance transcription of a downstream transgene and the accessibility of its promoter region to micrococcal nuclease. Plant Mol Biol **51,** 665-675.

Garrick, D., Fiering, S., Martin, D.I., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. Nat Genet **18,** 56-59.

Gasser, S.M., and Laemmli, U.K. (1987). A glimpse at chromosomal order. Trends Genet **3,** 16-22.

Gerdes, M.G., Carter, K.C., Moen, P.T., Jr., and Lawrence, J.B. (1994). Dynamic changes in the higher-level chromatin organization of specific sequences revealed by in situ hybridization to nuclear halos. J Cell Biol **126,** 289-304.

Glazko, G.V., Rogozin, I.B., and Glazkov, M.V. (2001). Comparative study and prediction of DNA fragments associated with various elements of the nuclear matrix. Biochim Biophys Acta **1517,** 351-364.

Gorbunova, V., and Levy, A.A. (1997). Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res **25,** 4650-4657.

Greger, I.H., and Proudfoot, N.J. (1998). Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. EMBO J **17,** 4771-4779.

Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science **301,** 798-802.

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

Han, K.H., Ma, C.P., and Strauss, S.H. (1997). Matrix attachment regions (MARs) enhance transformation frequency and transgene expression in poplar. Transgenic Res **6,** 415-420.

Henikoff, S. (1998). Conspiracy of silence among repeated transgenes. Bioessays **20,** 532-535.

Holmes-Davis, R., and Comai, L. (2002). The matrix attachment regions (MARs) associated with the Heat Shock Cognate 80 gene (HSC80) of tomato represent specific regulatory elements. Mol Genet Genomics **266,** 891-898.

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.

Huber, M.C., Bosch, F.X., Sippel, A.E., and Bonifer, C. (1994). Chromosomal position effects in chicken lysozyme gene transgenic mice are correlated with suppression of DNase I hypersensitive site formation. Nucleic Acids Res **22,** 4195-4201.

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

Ingelbrecht, I.L., Herman, L.M., Dekeyser, R.A., Van Montagu, M.C., and Depicker, A.G. (1989). Different 3' end regions strongly influence the level of gene expression in plant cells. Plant Cell **1,** 671-680.

Izaurralde, E., Kas, E., and Laemmli, U.K. (1989). Highly preferential nucleation of histone H1 assembly on scaffold-associated regions. J Mol Biol **210,** 573-585.

Jackson, D.A., and Cook, P.R. (1985). A general method for preparing chromatin containing intact DNA. EMBO J **4,** 913-918.

Jarman, A.P., and Higgs, D.R. (1988). Nuclear scaffold attachment sites in the human globin gene complexes. EMBO J **7,** 3337-3344.

Johnson, K.R., Disney, J.E., Wyatt, C.R., and Reeves, R. (1990). Expression of mRNAs encoding mammalian chromosomal proteins HMG-I and HMG-Y during cellular proliferation. Exp Cell Res **187,** 69-76.

Kas, E., and Chasin, L.A. (1987). Anchorage of the Chinese hamster dihydrofolate reductase gene to the nuclear scaffold occurs in an intragenic region. J Mol Biol **198,** 677-692.

Kas, E., Poljak, L., Adachi, Y., and Laemmli, U.K. (1993). A model for chromatin opening: stimulation of topoisomerase II and restriction enzyme cleavage of chromatin by distamycin. EMBO J **12,** 115-126.

Kay, V., and Bode, J. (1994). Binding specificity of a nuclear scaffold: supercoiled, single-stranded, and scaffold-attached-region DNA. Biochemistry **33,** 367-374.

Kipp, M., Gohring, F., Ostendorp, T., van Drunen, C.M., van Driel, R., Przybylski, M., and Fackelmayer, F.O. (2000). SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. Mol Cell Biol **20,** 7480-7489.

Kohli, A., Twyman, R.M., Abranches, R., Wegel, E., Stoger, E., and Christou, P. (2003). Transgene integration, organization and interaction in plants. Plant Mol Biol **52,** 247-258.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science **184,** 868-871.

Lechtenberg, B., Schubert, D., Forsbach, A., Gils, M., and Schmidt, R. (2003). Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. Plant J **34,** 507-517.

Levee, V., Garin, E., Klimaszewska, K., and Seguin, A. (1999). Stable genetic transformation of white pine (Pinus strobus L.) after cocultivation of embryogenic tissues with Agrobacterium tumefaciens. Mol Breeding **5,** 429-440.

Liebich, I., Bode, J., Frisch, M., and Wingender, E. (2002). S/MARt DB: a database on scaffold/matrix attached regions. Nucleic Acids Res **30,** 372-374.

Liu, J.W., and Tabe, L.M. (1998). The influences of two plant nuclear matrix attachment regions (MARs) on gene expression in transgenic plants. Plant Cell Physiol **39,** 115-123.

Makarevitch, I., Svitashev, S.K., and Somers, D.A. (2003). Complete sequence analysis of transgene loci from plants transformed via microprojectile bombardment. Plant Mol Biol **52,** 421-432.

Mankin, S.L., Allen, G.C., Phelan, T., Spiker, S., and Thompson, W.F. (2003). Elevation of transgene expression level by flanking matrix attachment regions (MAR) is promoter dependent: a study of the interactions of six promoters with the RB7 3' MAR. Transgenic Res **12,** 3-12.

Marsden, M.P., and Laemmli, U.K. (1979). Metaphase chromosome structure: evidence for a radial loop model. Cell **17,** 849-858.

Martelli, A.M., Falcieri, E., Zweyer, M., Bortul, R., Tabellini, G., Cappellini, A., Cocco, L., and Manzoli, L. (2002). The controversial nuclear matrix: a balanced point of view. Histol Histopathol **17,** 1193-1205.

Martens, J.H., Verlaan, M., Kalkhoven, E., Dorsman, J.C., and Zantema, A. (2002). Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. Mol Cell Biol **22,** 2598-2606.

Matzke, M., and Matzke, A.J. (2003). RNAi extends its reach. Science **301,** 1060-1061.

Matzke, M.A., Mette, M.F., and Matzke, A.J. (2000). Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant Mol Biol **43,** 401-415.

Maximova, S., Miller, C., Antunez de Mayolo, G., Pishak, S., Young, A., and Guiltinan, M.J. (2003). Stable transformation of Theobroma cacao L. and influence of matrix attachment regions on GFP expression. Plant Cell Rep **21,** 872-883.

Mendu, N., Massel, M., and Spiker, S. (2001). Increasing loop domain size does not diminish effects of matrix attachment regions on transgene expression in tobacco cells culture. Febs Lett **496,** 66-67.

Mette, M.F., van der Winden, J., Matzke, M.A., and Matzke, A.J. (1999). Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans. EMBO J **18,** 241-248.

Michalowski, S.M., Allen, G.C., Hall, G.E., Jr., Thompson, W.F., and Spiker, S. (1999). Characterization of randomly-obtained matrix attachment regions (MARs) from higher plants. Biochemistry **38,** 12795-12804.

Mirkovitch, J., Mirault, M.E., and Laemmli, U.K. (1984). Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell **39,** 223-232.

Mlynarova, L., and Nap, J.P. (2003). A self-excising Cre recombinase allows efficient recombination of multiple ectopic heterospecific lox sites in transgenic tobacco. Transgenic Res **12,** 45-57.

Mlynarova, L., Hricova, A., Loonen, A., and Nap, J.P. (2003). The presence of a chromatin boundary appears to shield a transgene in tobacco from RNA silencing. Plant Cell **15,** 2203-2217.

Mlynarova, L., Jansen, R.C., Conner, A.J., Stiekema, W.J., and Nap, J.P. (1995). The MAR-mediated reduction in position effect can be uncoupled from copy number-dependent expression in transgenic plants. Plant Cell **7,** 599-609.

Mlynarova, L., Loonen, A., Mietkiewska, E., Jansen, R.C., and Nap, J.P. (2002). Assembly of two transgenes in an artificial chromatin domain gives highly coordinated expression in tobacco. Genetics **160,** 727-740.

Mlynarova, 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.

Muskens, M.W.M., Vissers, A.P.A., Mol, J.N.M., and Kooter, J.M. (2000). Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing. Plant Mol Biol **43,** 243-260.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. Plant Cell **2,** 279-289.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science **183,** 330-332.

Padidam, M., and Cao, Y. (2001). Elimination of transcriptional interference between tandem genes in plant cells. Biotechniques **31,** 328-330, 332-324.

Paulson, J.R., and Laemmli, U.K. (1977). The structure of histone-depleted metaphase chromosomes. Cell **12,** 817-828.

Pawlowski, W.P., and Somers, D.A. (1996). Transgene inheritance in plants genetically engineered by microprojectile bombardment. Mol Biotechnol **6,** 17-30.

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

Pederson, T. (1998). Thinking about a nuclear matrix. J Mol Biol **277,** 147-159.

Petersen, K., Leah, R., Knudsen, S., and Cameron-Mills, V. (2002). Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. Plant Mol Biol **49,** 45-58.

Phi-Van, L., and Stratling, W.H. (1996). Dissection of the ability of the chicken lysozyme gene 5' matrix attachment region to stimulate transgene expression and to dampen position effects. Biochemistry **35,** 10735-10742.

Phi-Van, L., von Kries, J.P., Ostertag, W., and Stratling, W.H. (1990). The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. Mol Cell Biol **10,** 2302-2307.

Pikaard, C.S. (1998). Chromosome topology-organizing genes by loops and bounds. Plant Cell **10,** 1229-1232.

Romig, H., Ruff, J., Fackelmayer, F.O., Patil, M.S., and Richter, A. (1994). Characterization of 2 intronic nuclear-matrix-attachment regions in the human DNA topoisomerase-I gene. Eur J Biochem **221,** 411-419.

Ronai, D., Berru, M., and Shulman, M.J. (1999). Variegated expression of the endogenous immunoglobulin heavy-chain gene in the absence of the intronic locus control region. Mol Cell Biol **19,** 7031-7040.

Salomon, S., and Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J **17,** 6086-6095.

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

Schramke, V., and Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science **301,** 1069-1074.

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

Sidorenko, L., Bruce, W., Maddock, S., Tagliani, L., Li, X., Daniels, M., and Peterson, T. (2003). Functional analysis of two matrix attachment region (MAR) elements in transgenic maize plants. Transgenic Res **12,** 137-154.

Singh, G.B., Kramer, J.A., and Krawetz, S.A. (1997). Mathematical model to predict regions of chromatin attachment to the nuclear matrix. Nucleic Acids Res **25,** 1419-1425.

Somers, D.A., and Makarevitch, I. (2004). Transgene integration in plants: poking or patching holes in promiscuous genomes? Curr Opin Biotechnol **15,** 126-131.

Stief, A., Winter, D.M., Stratling, W.H., and Sippel, A.E. (1989). A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature **341,** 343-345.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H., and Martienssen, R. (1995). Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev **9,** 1797-1810.

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

Tan, J.H., Wooley, J.C., and LeStourgeon, W.M. (2000). Nuclear matrix-like filaments and fibrogranular complexes form through the rearrangement of specific nuclear ribonucleoproteins. Mol Biol Cell **11,** 1547-1554.

Tanabe, H., Muller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., Solovei, I., Cremer, C., and Cremer, T. (2002). Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc Natl Acad Sci U S A **99,** 4424-4429.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol **83,** 403-427.

Thompson, A.J., and Myatt, S.C. (1997). Tetracycline-dependent activation of an upstream promoter reveals transcriptional interference between tandem genes within T-DNA in tomato. Plant Mol Biol **34,** 687-692.

Ülker, B., Allen, G.C., Thompson, W.F., Spiker, S., and Weissinger, A.K. (1999). A tobacco matrix attachment region reduces the loss of transgene expression in the progeny of transgenic tobacco plants. Plant J **18,** 253-263.

Vain, P., James, A., Worland, B., and Snape, W. (2002). Transgene behaviour across two generations in a large random population of transgenic rice plants produced by particle bombardment. Theor Appl Genet **105,** 878-889.

Vain, P., Worland, B., Kohli, A., Snape, J.W., Christou, P., Allen, G.C., and Thompson, W.F. (1999). Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J **18,** 233-242.

Valerius, O., Brendel, C., Duvel, K., and Braus, G.H. (2002). Multiple factors prevent transcriptional interference at the yeast ARO4-HIS7 locus. J Biol Chem **277,** 21440-21445.

van der Geest, A.H., and Hall, T.C. (1997). The beta-phaseolin 5' matrix attachment region acts as an enhancer facilitator. Plant Mol Biol **33,** 553-557.

van der Geest, A.H.M., Hall, G.E., Spiker, S., and Hall, T.C. (1994). The beta-phaseolin gene is flanked by matrix attachment regions. Plant J **6,** 413-423.

van Leeuwen, W., Mlynarova, L., Nap, J.P., van der Plas, L.H., and van der Krol, A.R. (2001). The effect of MAR elements on variation in spatial and temporal regulation of transgene expression. Plant Mol Biol **47,** 543-554.

Vaucheret, H. (1993). Identification of a general silencer for 19s and 35s promoters in a transgenic tobacco plant - 90 bp of homology in the promoter sequence are sufficient for trans-inactivation. Comptes Rendus Acad. Sci. Ser. III-Sci. Vie-Life Sci. **316,** 1471-1483.

Vaucheret, H., Elmayan, T., Thierry, D., van der Geest, A., Hall, T., Conner, A.J., Mlynarova, L., and Nap, J.P. (1998). Flank matrix attachment regions (MARs) from chicken, bean, yeast or tobacco do not prevent homology-dependent trans-silencing in transgenic tobacco plants. Mol Gen Genet **259,** 388-392.

Vergunst, A.C., and Hooykaas, P.J.J. (1999). Recombination in the plant genome and its application in biotechnology. Crit Rev Plant Sci **18,** 1-31.

Villemure, J.F., Savard, N., and Belmaaza, A. (2001). Promoter suppression in cultured mammalian cells can be blocked by the chicken beta-globin chromatin insulator 5'HS4 and matrix/scaffold attachment regions. J Mol Biol **312,** 963-974.

Wei, X., Somanathan, S., Samarabandu, J., and Berezney, R. (1999). Three-dimensional visualization of transcription sites and their association with splicing factor-rich nuclear speckles. J Cell Biol **146,** 543-558.

Weiler, K.S., and Wakimoto, B.T. (1995). Heterochromatin and gene expression in Drosophila. Annu Rev Genet **29,** 577-605.

Wolffe, A.P., and Matzke, M.A. (1999). Epigenetics: regulation through repression. Science **286,** 481-486.

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. Nature **419,** 641-645.

Ye, F., and Signer, E.R. (1996). RIGS (repeat-induced gene silencing) in Arabidopsis is transcriptional and alters chromatin configuration. Proc Natl Acad Sci U S A **93,** 10881-10886.

Zhao, K., Kas, E., Gonzalez, E., and Laemmli, U.K. (1993). SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. EMBO J **12,** 3237-3247.

CHAPTER 2

Quantification of Green Fluorescent Protein Expression in Plant Cells by Flow Cytometry

Introduction

Throughout the course of basic and applied plant molecular biology, researchers often want to know how genes are regulated both temporally and spatially. That is, when and where are genes turned on and off? An important factor in answering these questions is the availability of markers or reporter genes to be used in transgenic studies. One of the most commonly used reporter genes is the *uidA* gene from *Escherichia coli* that encodes the enzyme β-glucuronidase (GUS) (Jefferson et al., 1987). This gene can be fused to endogenous promoters or other regulatory elements and used to track when and where GUS is produced. Measurement of GUS requires the addition of enzyme substrates that are cleaved into products with either visible or fluorescent properties. One of the great advantages of using enzymes as reporter genes is signal amplification, given that every enzyme molecule expressed is capable of producing many detectable products. As an example, advances in our understanding of flower development have benefited greatly from the use of *uidA* gene by fusing regulatory elements involved in regulating patterns of expression in specific tissues to give rise to the classic ABC model (Sieburth and Meyerowitz, 1997).

On of the major drawbacks for the analysis of GUS expression *in vivo* is that plant tissues need to be fixed in order for substrates to penetrate evenly. While this invasive

process gives qualitative information, quantitative analysis has been much more difficult to achieve due to factors such as: substrate access, differential retention between substrates and products, pH microenvironments, and differential product accumulation within subcellular compartments (Galbraith et al., 1995). Given these difficulties, quantitative measurements are most often performed on tissue extracts which homogenize GUS expression patterns, thus losing this spatial information.

Green fluorescent protein (GFP) can be monitored without tissue destruction or the need for exogenously added substrates. GFP is an intrinsically fluorescent 26 kDa protein originally cloned from the jellyfish *Aequorea victoria* (Prasher et al., 1992). Both prokaryotic and eukaryotic organisms are able to express GFP and retain its green fluorescence properties (Chalfie et al., 1994). Formation of the GFP chromophore involves a series of posttranslational intramolecular reactions of amino acids 65-67 (Ser-Tyr-Gly) to form a *p*-hydroxybenzylidene-imidazolidinone species. This wild-type chromophore is excited by either UV or blue light (maximally at 396 nm or 475 nm) and emits green fluorescence maximally at 508 nm (Heim et al., 1994).

Amino acid substitutions at the chromophore core can alter excitation and/or emission wavelengths to produce specialized GFP variants. See Galbraith et al. (1999b) for a review. Of particular interest are variants that alter the excitation properties of GFP to coincide with the 488 nm argon laser commonly used in fluorescence microscopy and flow cytometry. "Red-shifted" variants are excited maximally near 475 nm, caused by a serine to threonine substitution at amino acid position 65 and have sevenfold higher fluorescence (Heim et al., 1995). These variants are called "red-shifted," since wild-type

GFP has two excitation maxima (396 nm and 475 nm) and these variants have only the longer wavelength 475 nm maximum (closer or "shifted" toward the red end of the visible spectrum).

Besides changes in the fluorescent properties, other modifications were necessary to fully exploit the potential of GFP as a reporter gene in plants. Aberrant GFP mRNA splicing was observed in *Arabidopsis thaliana* and corrected by altering codons around splice junctions (Haseloff et al., 1997). Even though a protein is expressed efficiently, it may not be functional. For example, when GFP aggregates, its fluorescent properties greatly diminish. By altering codons to increase the solubility of GFP, Crameri et al. (1996) increased the total amount of fluorescence emitted from cells. Finally, Davis and Vierstra (1998) combined all of these strategies to create a codon optimized soluble modified red-shifted GFP (smRS-GFP) variant suitable for use in higher plants. This variant was used this study.

Flow cytometry is a technique used to measure physical or chemical properties of cells or particles as they move in a fluid stream (Figure 2-1). The basic principal, when used to study cells, is that single-cell suspensions pass in single file through a flow chamber, passing through a laser beam. As each cell enters the path of the laser beam, light is scattered in all directions by diffraction, refraction, and reflection. The amount and directions of light scatter are influenced by physical characteristics of the particle. All flow cytometers measure two types of light scatter: forward-angle light scatter (FSC) and side-angle light scatter (SSC). Forward-scattered light is collected at 0.5-10º angles relative to the incident beam. Light scatter collected by FSC is the result of diffraction.

Larger particles diffract more light, consequently FSC can be used to estimate particle size. Side-scattered light is collected at a 90º angle relative to the incident beam. Light scatter collected by SSC is the result of refraction and reflection. Particles with greater surface/membrane irregularities or internal granularity tend to scatter more light toward the SSC detector. Using measurements of FSC and SSC, subpopulations of particles can often be discriminated based on their physical properties (Jaroszeski and Heller, 1998). This technique is very useful, as cell preparations are rarely pure and often contain cellular debris. In the case of the work reported in this dissertation, both debris and undissociated clusters of tobacco protoplasts can be distinguished by their size and granularity as determined from light scattering measurements.

In addition to FSC and SSC measurements, various wavelengths of fluorescence can be detected. Most modern cytometers are equipped with an argon laser that emits light at 488 nm and three detectors that are capable of collecting fluorescence at three narrow ranges of independent wavelengths called channels. While narrow ranges can be used to detect wavelengths of maximal emission from fluorochromes and fluorophores, unfortunately these fluorescent molecules do not emit a single wavelength of light, but rather a spectrum of wavelengths that can overlap with more than one detection range. This becomes a problem only when more than one fluorescent molecule is measured. For example, the emission spectrum of GFP has a maximum at 508 nm and is detected in the 510-540 nm green fluorescence channel. Propidium iodide has an emission maximum at 617 nm and is detected in the 564-606 yellow/orange fluorescence channel. See figure 2-2. The emission spectrum of GFP slopes downward toward the point at which a portion

of the long wavelength tail overlaps with yellow/orange detection. This spectral overlap would result in an overestimation of the amount of propidium iodide staining in samples with green fluorescence. Fortunately, this problem can be alleviated by adjusting compensation settings between the two detection ranges to account for the spectral overlap (Shapiro, 1995). In other words, a portion of yellow/orange fluorescence is subtracted out depending on the levels of green fluorescence detected.

Once light scatter and fluorescence measurements have been collected for each particle, they are stored digitally in a list-mode file. This file stores all of the acquired raw data from every particle that goes through the flow cytometer. Several specialized software packages are available to analyze and display the large amounts of data in list-mode files. For the purposes of the work reported in this dissertation, CellQuest 3.3 from Becton-Dickinson was used to analyze flow cytometry data obtained on a FACSCalibur cytometer, also from Becton-Dickinson. A more in depth discussion of data analysis will follow as it pertains to the work presented here.

Methods and Results

Plasmids

In order to develop methods to measure GFP expression in tobacco suspension culture cells, the soluble-modified, red-shifted GFP with its enhanced fluorescence (Davis and Vierstra, 1998) was used. As mentioned previously, the red-shifted variant has a unimodal excitation peak at 475 nm, which is near the 488 nm laser used in most flow cytometers. A binary vector (pGFP) containing this GFP cassette as well as the

nptII kanamycin resistance gene cassette and binary vector backbone from pGPTV-KAN (Becker et al., 1992) was kindly provided by Luke Mankin and originally described as pLMNC-b47 (Mankin, 2000). Briefly, transcription of the soluble-modified, red-shifted green fluorescent protein gene is driven by the 35S promoter of CaMV firing away from the right T-DNA border and terminated by the nos terminator region. Transcription of the neomycin phosophotransferase selectable marker gene (NPTII) is driven by the nos promoter firing away from the GFP cassette toward the left T-DNA border and terminated by the g7T polyadenylation region. See figure 3-1. The pGFP binary vector was electroporated into an *Agrobacterium tumefaciens* strain (UIA143/ pMP90) described by (Callaway et al., 1996).

Plant Material and Agrobacterium-mediated Transformation

A tobacco suspension culture stock cell line (NT-1) was maintained at 27ºC under constant light conditions on an orbital shaker (125 rpm). Weekly passages of 3% of the total volume of the culture were conducted (3 mL into 100 mL of fresh liquid media containing the following: 1X MS salts, 3.0 mL/L Miller's solution I (60 g/L KH_2PO_4), 10 mL/L B1/inositol vitamins (10 g/L myoInositol, 0.1 g/L thiamine HCl), 200 µL/L 2,4-D (1 mg/mL^{EtOH}) with the pH adjusted to 5.7 with 1 M KOH). *Agrobacterium*-mediated transformation of NT-1 cells was performed as described by An (1985). Briefly, 4 day old NT-1 liquid culture cells (4 mL) were infected with an *Agrobacterium tumefaciens* strain (100 µL, $OD_{600} \approx 1.0$) containing pGFP in a 100 mm petri plate and placed in the dark at room temperature. After 48 hours, cells were resuspended in 4 mL of fresh media

and 400 μL of this suspension was plated on solid media containing 200 mg/L Timintin and 50 mg/L kanamycin. After 2-4 weeks, microcalli appeared as small colonies rising above the lawn of untransformed NT-1 cells. Microcalli were transferred to a second plate of fresh medium. After two weeks of growth (in which the calli reached approximately 10 mm in diameter), a small, randomly selected, piece of the callus tissue (approximately 1 mm in diameter) was transferred to a third plate of fresh medium. After two additional weeks, entire calli from a randomly selected subset of calli were dispersed independently in 5mL aliquots of fresh liquid media supplemented with 200 mg/L Timintin and 50 mg/L kanamycin. Cells in liquid cultures were transferred weekly by adding 0.5 mL to 5 mL of fresh media and grown under the same conditions as previously stated for the NT-1 stock culture.

Protoplast Preparation

Single-cell suspensions are necessary to allow passage through the 100 μm orifice of the flow cell to collect data on individual cells. Because tobacco suspension cells grow in filamentous structures that cannot be analyzed by flow cytometry, it is necessary to disrupt the cell walls that hold these filaments together. A gentle digestion with cellulases and pectolyases releases individual cells as protoplasts that can pass one at a time through the flow cell of the cytometer.

Protoplasts were obtained from four day old NT-1 liquid cultures by the following procedure. A protoplasting solution was composed of 0.4 M manitol, 10 mM MES pH 5.7, 1 mg/mL Pectolyase Y-23 and 10 mg/mL Cellulase RS (Karlan Research Products

Corporation, Santa Rosa, CA). Pelleted cells (approximately a 300 μL packed cell volume) were washed by suspending in a 0.4 M manitol, 10 mM MES pH 5.7 solution, pelleting at 450 g for 3 minutes, and resuspending in 500 μL of protoplasting solution. Incubation for 45 minutes at 28ºC on a rotating shaker (125 rpm) yielded greater than 75% liberated protoplasts. Longer incubation times were avoided, as they tended to result in more free nuclei and cellular debris from ruptured protoplasts. It was not necessary to liberate 100% of the cells as protoplasts since the final step in the protoplasting procedure was passage through a 50 μm mesh; any remaining filaments were filtered out of the final preparation. The remaining steps were carried out at 4ºC. Protoplasts were centrifuged for 3 minutes at 100 g, washed in a 0.4 M manitol, 10 mM MES pH 5.7 solution, and resuspended in 1 mL of 0.4 M manitol, 10 mM MES solution. G-forces greater than 100 g resulted in fewer intact protoplasts. Samples were then allowed to pass through a 50 μm mesh and stored on ice up to two hours prior to flow cytometry analysis.

Flow Cytometry: Initial Setup

The FACSCalibur (Becton-Dickinson, San Jose, CA) flow cytometer at North Carolina State University's College of Veterinary Medicine's Flow Cytometry Facility was the primary instrumentation used in this work. The FACSCalibur is equipped with a 15 mW 488 nm, air-cooled argon-ion laser, FSC and SSC detectors, and four fluorescence detectors and bandpass filters with the following ranges: FL1 (515-545 nm), FL2 (564-606 nm), FL3 (653-669 nm), and FL4 (>670 nm). The FSC detector was

operated at a gain setting of 9.9 and a high voltage of E-1 V. The SSC detector high voltage was set to 319 V with no gain adjustments. The PMT high voltages were 370-450 V (green channel) based on standardizing fluorescence to calibration beads (discussed in the next section) and 447 V (yellow/orange channel) with no gain adjustments for either. Particles passed through a sort-sense flow cell equipped with a 100 nm orifice.

Flow Cytometer Calibration of Green Fluorescence

To insure degradation of laser intensity and other instrument-specific fluctuations did not affect our results taken over several time points, Flow Check YG 6.0 microspheres (Polysciences Inc., Warrington, PA) were used to monitor and calibrate the flow cytometer. Microspheres with three distinct intensities of green fluorescence provided three calibration points (Figure 2-3A) to adjust the PMT voltage of the green channel. Voltage adjustments were made such that microspheres gave the following approximate green fluorescence relative units: high intensity (3,500 RFU), intermediate intensity (680 RFU), and low intensity (45 RFU). These calibration points were used for all experiments until a new microsphere calibration set was needed. Because the fluorescence intensities of the microspheres in one set may not match those of a second set from the same manufacturer, the PMT voltage of the green channel was adjusted as previously described for the original set, and then with the same instrument settings, microspheres from the second set were analyzed (Figure 2-3B). The following green

fluorescence values became the new calibration points used for subsequent analyses: high intensity (5,695 RFU), intermediate intensity (711 RFU), and low intensity (43 RFU).

Green Fluorescent Protein Quantitation in NT-1 Cells using Light Scatter Gating

Protoplasts prepared from NT-1 cultures were subjected to flow cytometric analysis. Initially data from 30,000 particles (events) were collected from several individual cell lines. Data from an untransformed, wild-type cell line is plotted in figure 2-4. A biparametric anaylisis of light scatter properties of protoplast preparations is represented in figure 2-4A. A large cluster of particles with intermediate SSC values overlaps with a smaller cluster of lower SSC values. Particles with very low SSC values were not collected by defining an "acquisition gate." This gate sets a threshold value of SSC the particle must have, 15 SSC light units in this example, in order for any data to be collected. (Note the region with SSC values <15 does not have any data points in figure 2-4, panels A and B.) In addition to having very low SSC (surface granularity) values, these excluded particles also had very low FSC (size) values (data not shown), suggesting these particles may be nuclei or other intracellular debris that would otherwise overwhelm list-mode data files.

Figure 2-4B illustrates the detection of green fluorescence in all particles (with SSC values above 15) from a protoplast preparation of wild-type NT-1 cells. Many of the particles have low levels of intrinsic green fluorescence presumably from low levels of autofluorescence. NT-1 tobacco suspension culture cells are inherently non-photosynthetic. They do not contain enough chlorophyll to make them noticeably green

and unlike photosynthetic tissues, they appear light yellow in color. The observed autofluorescence is most likely the result of other endogenous pigments as previously observed in etiolated tobacco protoplasts analyzed by flow cytometry (Galbraith et al., 1999a). These low levels of green fluorescence actually prove useful to distinguish untransformed cells from cellular debris that tend to have nearly undetectable levels of green fluorescence. This debris can be seen in figure 2-4B as a region containing a large number of particles with nearly undetectable levels of green fluorescence in the range of 10^0, immediately adjacent to the y-axis, with low SSC values, below 75 light units. Cells with damaged membranes often have decreased SSC values as McGann et al. (1988) discovered when they osmotically stressed and freeze-thawed cells. A clearer picture of this debris can be seen in a histogram plot of green fluorescence (Figure 2-4C). The number of counts on the y-axis is off scale with nearly undetectable levels of green fluorescence (Figure 2-4C). This can be seen in this graph only by a thickening of the y-axis (compare the y-axis of panel C with panel D in figure 2-4). If these debris particles were included in our analysis, they would lead to an overestimate of the proportion of protoplasts that do not express GFP.

"Bitmap gating" is a method that can be used to exclude these particles with low levels of side scatter and nearly undetectable levels of green fluorescence. This process is similar to "acquisition gating" discussed earlier, in that a range of measurements are defined and are either included or excluded from the analysis. While the acquisition gate defines the particles from which data will be collected, a bitmap gate defines the portion of the collected data that will be used for further analysis. Future references to gating

will refer to "bitmap gating" unless otherwise specified. By defining a gated region around the uppermost cluster in figure 2-4A labeled "R1," particles with low levels of side scatter can be excluded from further analysis (14% in this example). This is shown in figure 2-4D where the histogram is composed only of particles from the gated region. Notice that many of the data points with nearly undetectable levels of green fluorescence in the range of 10^0 are excluded (compare the "width" of the y-axis of panel C with panel D in figure 2-4).

While the gating approach based only on light scattering appeared to be successful in protoplasts prepared from wild-type NT-1 cells, its limitations became evident when protoplasts from transgenic NT-1 cell lines carrying pGFP were analyzed. Figure 2-5A displays light scatter properties of one such line and figure 2-5B demonstrates an approximate 100 fold increase in green fluorescence exhibited by a subset of the transgenic protoplasts. Similar to our analysis of wild-type protoplasts, all green fluorescence data were plotted (Figure 2-5C) or only gated events from figure 2-5A were plotted (Figure 2-5D). It also appears that a substantial portion of cells have low levels of green fluorescence similar to those of wild-type cells (Figure 2-5D). This result contrasts to our microscopic observation that prior to protoplasting all cells of this cell line exhibited uniform, non-variegated green fluorescence (data not shown). It was not clear what type of particles gave rise to the histogram peak consistent with wild-type levels of green autofluorescence, but such particles would interfere with our ability to accurately quantitate GFP expression in individual cells.

Another observation that lends evidence to the fact that gating based on light scatter properties alone may not be sufficient is that occasionally cells with different levels of green fluorescence also had different physical properties. Analysis of a transgenic pGFP cell line is represented in figure 2-6. Gating based solely on light scatter properties (Figure 2-6A) resulted in a histogram with three modes of green fluorescence (Figure 2-6B). Observations of more than one mode of green fluorescence were not uncommon, but what was unusual about this cell line and others was that the intermediate mode of green fluorescence had increased levels of forward angle light scatter (Figure 2-6C). Since FSC is a measure of particle size, why should intermediate levels of green fluorescence be associated with greater FSC? One possible explanation is that these protoplasts are damaged in a way that would cause enlargement and some diffusion of GFP out of the protoplast. Several observations lead us to believe that this may be the case. Closer examination of data from all other cell lines that exhibited a similar pattern of increased FSC revealed that in every case where a mode of green fluorescence had increased FSC, there was always a mode with greater green fluorescence. In other words, there was no example of the highest mode of green fluorescence ever having an increase in FSC compared to the other modes. These observations suggested that intermediate modes of green fluorescence may have "originated" from protoplasts with higher levels of GFP expression. That is, protoplasts that originally had higher levels of GFP may be damaged in such a way that a portion of soluble GFP diffuses out into the medium giving rise to an additional mode with reduced green fluorescence. These damaged protoplasts may have higher FSC because they may be osmoticly stressed resulting in a larger cell

volume. If GFP were actually diffusing out of protoplasts, then total green fluorescence would be underestimated. This possibility was further investigated using propidium idodide based gating.

Green Fluorescent Protein Quantitation in NT-1 Cells using Propidium Iodide Gating

In the process of generating protoplasts from intact NT-1 filaments, nuclei and cellular debris can be expected to be produced. Additionally, protoplasts with compromised plasma membrane integrity may also be present in protoplast preparations. In our analysis above, it appears that debris and compromised protoplasts may interfere with accurately quantitating GFP expression. One method that has been commonly employed with animal cells is to use a dye exclusion test (Shapiro, 1995). A DNA stain such as propidium iodide is unable to penetrate intact plasma membranes, but readily enters the cell to stain the nucleus when a breach in the plasma membrane occurs. Not only would such a breach allow propidium iodide into the cell, but soluable GFP would diffuse out into the medium.

In order to illustrate these phenomena, another cell line was identified that had high levels of uniform, non-variegated, GFP fluorescence prior to protoplasting. Microscopically, all cells in this cell line had levels of GFP fluorescence that were indistinguishable from one another (data not shown). Protoplasts were prepared from this cell line, stained with propidium iodide, and subjected to confocal microscopy (Figure 2-7). Clearly, GFP fluorescence is no longer uniform for all protoplasts. Protoplasts with compromised plasma membrane integrity allow GFP to diffuse out of the cell and

propidium iodide to diffuse in and stain the nucleus. Despite having defects in plasma membrane integrity, these protoplasts may still be able to maintain their size and general shape. Spurious flow cytometry data from these protoplasts and from free nuclei can be excluded from analysis by measuring and accounting for the high levels orange fluorescence associated with propidium iodide staining.

In order to evaluate the use of propidium iodide in our flow cytometry experiments, protoplasts from the same transgenic line as in figure 2-7 were analyzed (Figure 2-8). When gating based only on light scattering is used (Figure 2-8A), it appears as though a substantial proportion of the cells (32% in this example) have wild-type levels of green fluorescence (Figure 2-8B). Because microscopic evaluation of the cells prior to protoplasting showed that all cells displayed uniform GFP fluorescence, it appears that the protoplasts with compromised membranes and residual nuclei not gated away by light scattering cause an overestimation of the proportion of cells that have wild-type levels of green fluorescence. This problem can be alleviated by gating data based both on side light scatter (to exclude data from debris that is not consistent with light scatter properties of protoplasts) and on orange fluorescence associated with propidium iodide staining (to exclude data from free nuclei and protoplasts with compromised membranes that have allowed GFP to diffuse out of the cell and propidium iodide to diffuse in and stain nuclei). When the same data as shown in figure 2-8A and 2-8B were gated based on both side light scatter and orange fluorescence (Figure 2-8C), only 0.9% of the cells had fluorescence properties consistent with wild-type cells (Figure 2-8D).

It was not necessary to use propidium iodide based gating in all cases as is evidenced by figure 2-8E and 2-8F. A separate preparation of protoplasts from the same transgenic line as in figure 2-8A-D was subjected to flow cytometry analysis. In this case gating based on light scatter properties alone was sufficient to exclude all but 2.0% of particles with background levels of green fluorescence. Gating based on propidium iodide further reduced the percentage of particles with background fluorescence to 0.5% (data not shown). These results may be attributed to preparation-to-preparation variations in protoplast integrity that alter the effectiveness of gating based on light scatter properties alone. By gating based on background levels of propidium iodide staining, however, any effect of such protoplast preparations is removed from the final analysis.

Propidium iodide based gating was also used to determine the membrane integrity of protoplasts with intermediate modes of green fluorescence and unusually high FSC values that were previously observed (Figure 2-6). Gating based only on light scatter properties of a protoplast preparation from a similar transgenic cell line (Figure 2-9A) resulted in three modes of green fluorescence (Figure 2-9B). The intermediate mode has higher levels of forward light scatter than the highest green fluorescence mode (Figure 2-9C). When orange fluorescence from PI staining is plotted against green fluorescence, it is clear that the intermediate mode of green fluorescence also has high levels of orange fluorescence (Figure 2-9D). These results suggest the protoplasts in this intermediate mode have compromised plasma membrane integrity, in which case some soluble GFP may have diffused out of the cells. It's unknown why all GFP did not diffuse out of protoplasts in these cases. It's possible some GFP may be bound intracellularly or that a

breach in the plasma membrane happened fairly recently compared to the time of flow analysis and residual GFP has not had time to diffuse out. Gating based on PI staining (Figure 2-9E) again resulted in a unimodal green fluorescence histogram in this case (figure 2-9F). The position of the PI based gate was consistent with that of PI stained wild-type protoplasts (data not shown) and with the sample in figure 2-8. Particles with orange fluorescence greater than approximately 100 relative fluorescence units (RFUs) were inconsistent with intact wild-type protoplasts and were considered to be intracellular debris or compromised protoplasts. The intermediate green fluorescence mode in figure 2-9D fell outside this range and thus was not included in subsequent analysis.

Initially, we chose a propidium iodide concentration of 5 μg/mL. In order to determine whether lower concentrations of PI could give similar results, three different concentrations were compared (Figure 2-10). While lower concentrations did prove to be successful in gating away debris, the original concentration of 5 μg/mL gave the best orange fluorescence separation (compare figure 2-10A, C, and E). It's likely, that lower concentrations could have given similar results if the voltage from the FL2 detector (564-606 nm) was raised to increase the amount of PI fluorescence signal, but this approach seemed unnecessary since the cost associated with PI stain is negligible and there were no other drawbacks associated with the use of this level of PI. For all subsequent analysis, a concentration of 5 μg/mL of PI was used to stain protoplast preparations.

Since there is variation in protoplast preparation purity, collecting data from 30,000 total particles per sample will not standardize the number of viable protoplasts for which data was collected. Instead we choose an approach in which flow analysis was

continued until data from 15,000 viable protoplasts were collected. We used this gating approach in all analyses in which flow cytometry was used to assess the effect of MARs on transgene expression in chapter three.

Discussion

The use of GFP as a reporter in plants has become popular choice for plant biologists for its non-invasive detection, but quantitation of GFP expression *in vivo* still proves challenging. One approach is the use of fluorescence microscopy to quantitatively image GFP (Piston et al., 1999). We chose not to use this approach due to the large numbers of cell culture lines and the amount of time and expense required to collect data. Instead, we chose flow cytometry to quantitate GFP expression in individual tobacco cells in culture. While flow cytometry methods for the detection of GFP expression in plant cells have been previously published (Galbraith et al., 1995, 1999a,b; Doran, 2000; Bohanec et al., 2002), there still remained two complications that limited our ability to accurately quantitate GFP expression. Could levels of GFP expression be high enough to discriminate clusters of green fluorescence from autofluoresence? Could viable non-expressing protoplasts be distinguished from nonviable or compromised protoplasts?

The first problem was limited separation between wild-type green autofluorescence and green fluorescence associated with GFP expression. Figure 2-11 demonstrates this limited separation as evidenced by overlapping clusters of green fluorescence values between GFP expressing and non-expressing maize protoplasts transfected with a GFP-encoding construct (Figure 2-11A and B adapted from Galbraith

et al., 1995). Similar observations can be made with isolated nuclei from transgenic tobacco plants carrying a nuclear targeted red-shifted GFP variant (S65T) (Figure 2-11C and D adapted from Grebenok et al., 1997). In both of these examples of previously reported flow cytometry methods of detection of GFP expression, clusters of GFP expressing particles have green fluorescence that overlaps with non-expressing particles. This inability to completely resolve populations of GFP expressing and non-expressing protoplasts or nuclei could prove problematic in the quantification of GFP expression. In our work with protoplasts derived from transgenic NT-1 tobacco cells in culture, in many cases we observed distinct clusters of green fluorescence values that could be easily distinguished from those of wild-type protoplasts (Figure 2-11E and F and Figure 3-6B). It's not clear whether differences between our observations and those of Galbraith et al. (1995) and Grebenok et al. (1997) were due to less autofluorescence in NT-1 cells, the analysis of protoplasts isolated from stable transgenic lines compared to nuclei from transgenic plants, and/or the use of the soluble-modified, red-shifted GFP.

A much more serious problem was our initial inability to discriminate between intact protoplasts with no GFP expression and nonviable protoplasts, which may have lost the GFP they once contained. Protoplast preparations frequently contain free nuclei, cellular debris, and protoplasts with compromised membrane integrity. These entities may interfere with the quantitation of GFP expression in individual cells if they are not distinguished from viable cells. Galbraith et al. (1995) did not employ procedures to distinguish these entities from viable protoplasts and did not report any data regarding the purity of protoplast preparations. Similarly, Desikan et al. (1999) did not distinguish

between viable and non-viable rice protoplasts while performing flow cytometry analysis of GFP expression upregulated by abscisic acid (ABA). However, the authors were able to measure cell viability by the accumulation of green fluorescence from treatment with fluorescein diacetate. Cells which do not possess an intact cell membrane cannot accumulate the fluorescent product and therefore do not exhibit green fluorescence. The authors found increased concentrations of ABA reduced protoplast viability, but they did not simultaneously measure GFP expression and protoplast viability, since both require the detection of green fluorescence.

Initially, we tried to use light scatter properties to distinguish between viable cells and debris with limited success. Subsequently we employed a dye exclusion test whereby propidium iodide enters protoplasts to stain DNA containing debris only when the plasma membrane is breached. By gating to include only events that have background levels of orange fluorescence, we were able to quantitate GFP expression in individual cells with minimal side effects from debris. This technique has not been previously applied to flow cytometry of plant protoplasts, but has been frequently used in non-plant systems (e.g. Migliaccio et al., 2000).

Several applications of flow cytometry coupled with the GFP reporter could be implemented in plant molecular biology. One of the true powers of flow cytometry is analysis of heterogeneity in a cell population that may have resulted from epigenetic processes. One obvious application, which is the subject of the work presented in this dissertation, is to test various regulatory elements that may affect transgene expression. A recent application that has shown great promise is to put GFP under control of cell

specific promoters in whole plants, isolate protoplasts and use fluorescence activated cell sorting (FACS) to collect specific cell types (Birnbaum et al., 2003). In this manner the authors were able create cell type specific gene expression profiles of root development using microarrays. Alternatively, GFP could be used as a biological sensor with differentially regulated promoters. This was the case as previously mentioned when Desikan et al. (1999) monitored ABA signalling. This approach could be used to study other signalling cascades and, coupled with FACS, could allow for the isolation of cells at specific stages of the signalling pathway. Unfortunately, FACS is beyond the scope of this work, but the same techniques applied here may also prove valuable to flow sorting applications.

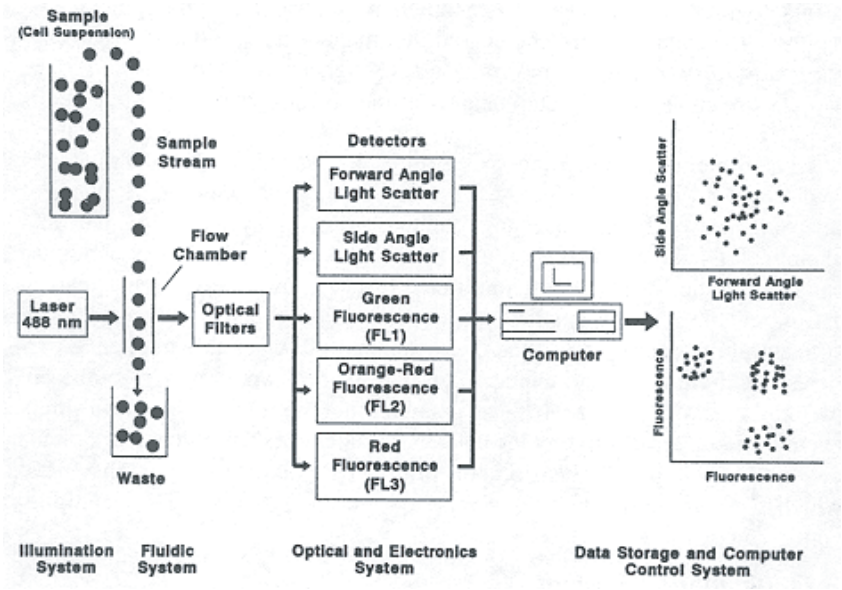


Figure 2-1. A Schematic of the Primary Components that Comprise a Flow Cytometer. Arrows indicate the flow of particles and information. A fluidic system transports particles or cells from a prepared suspension past a focused laser beam that is generated by an illumination system. Particle interrogation takes place, one cell at a time, in a flow chamber. The resulting scattered light and fluorescence is gathered by an optical and electronics system that translates the light signals into information that is saved by the data storage and computer control system. After data from a sample has been stored, retrospective graphical data analysis can be performed with the aid of software. (Reproduced from Jaroszeski and Heller, 1998)

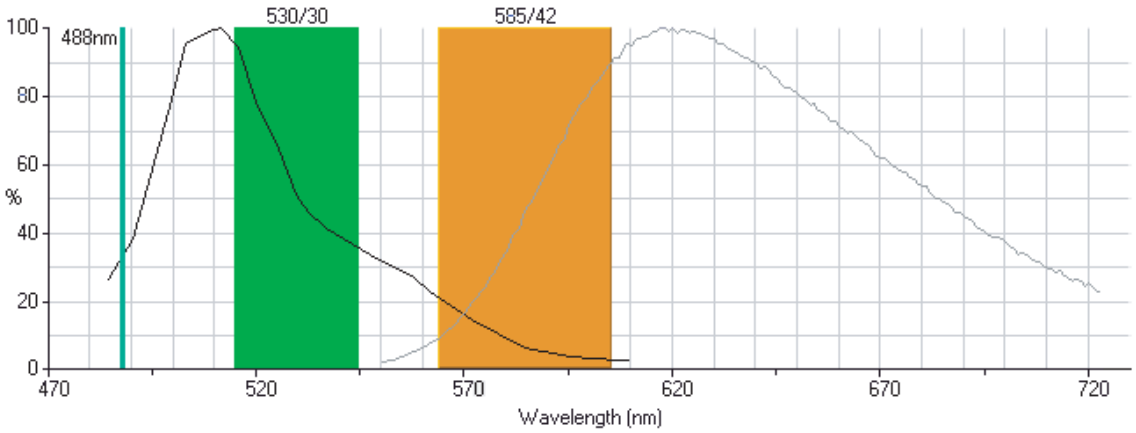


Figure 2-2. FACSCalibur Detection of Emission Spectra from Green Fluorescent Protein and Propidium Iodide.
Excitation of GFP and PI is provided by a 488 nm argon laser indicated by a green line. Emission spectra for GFP and PI are outlined in black and gray respectively. The 530 nm fluorescence detector (FL1, green shading) has a 30 nm bandpass filter to collect emission from GFP. The 585 nm fluorescence detector (FL2, orange shading) has a 42 nm bandpass filter, which primarily collects emission from PI, but also collects a small portion of emission from GFP due to the spectral overlap. This overlap is accounted for by adjusting compensation settings. (Spectra were obtained from Becton-Dickinson, http://www.bdbiosciences.com/spectra)

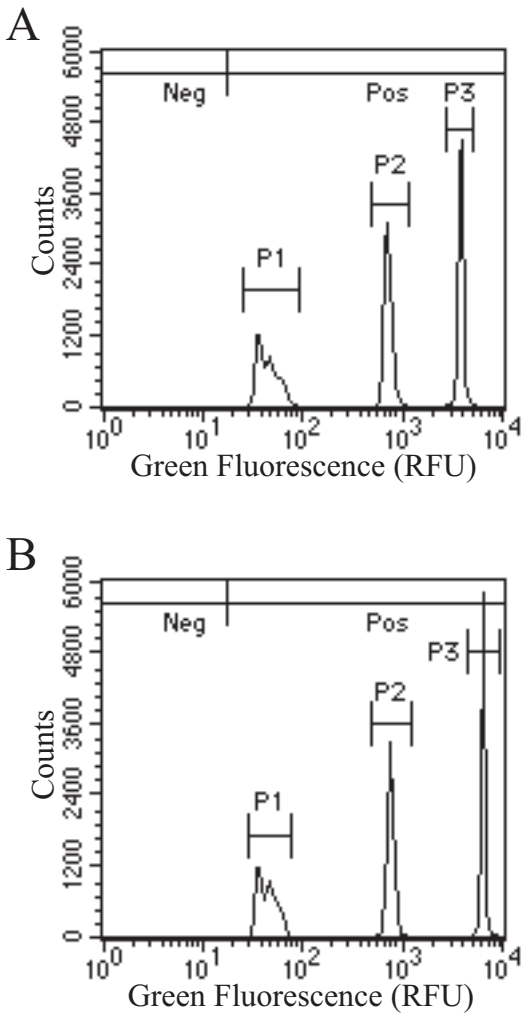


Figure 2-3. Green Fluorescence Histograms of Calibration Standards.
(A) Three independent calibration points were provided by flow cytometry analysis of Flow Check YG 6.0 microspheres from Polysciences Inc. (Warrington, PA). **(B)** A second set of microspheres had different intensities of green fluorescence compared to (A). Fluorescence was measured in Relative Fluorescence Units (RFUs). In subsequent histograms displaying fluorescence of cells, cells with green fluorescence below 17 RFUs are considered "GFP negative cells" (Neg) and cells with greater fluorescence are considered "GFP positive cells" (Pos). The bar at the top of the histograms contains a vertical line demarking the 17 RFU cutoff point between "Negative" (Neg) and "Positive" (Pos). Peak markers (P1, P2, and P3) denote peaks of interest to calculate additional histogram statistics.

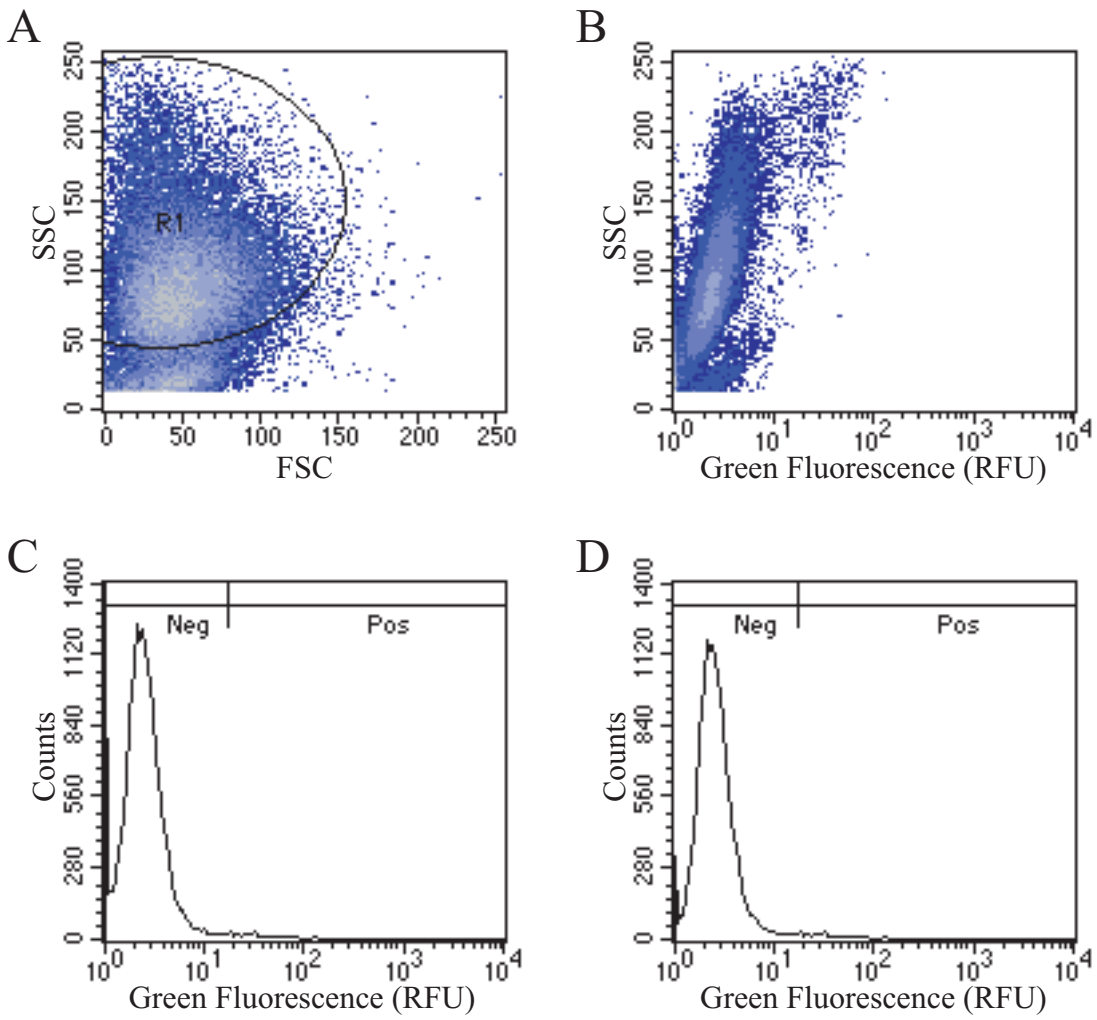


Figure 2-4. Green Fluorescence and Light Scatter Properties of Wild-type NT-1 Protoplast Preparations.
(A) Biparametric flow analysis of wild-type, untransformed control protoplasts, based on side light scatter (SSC) and forward light scatter (FSC). Gated events (R1) are plotted in (D). **(B)** Biparametric flow analysis of control protoplasts, based on side scatter and green fluorescence. **(C)** Green fluorescence histogram of all data from control protoplasts in (A). **(D)** Green fluorescence histogram of data from gated control protoplasts (R1) from (A). Parameters for histograms are the same as in figure 2-3.

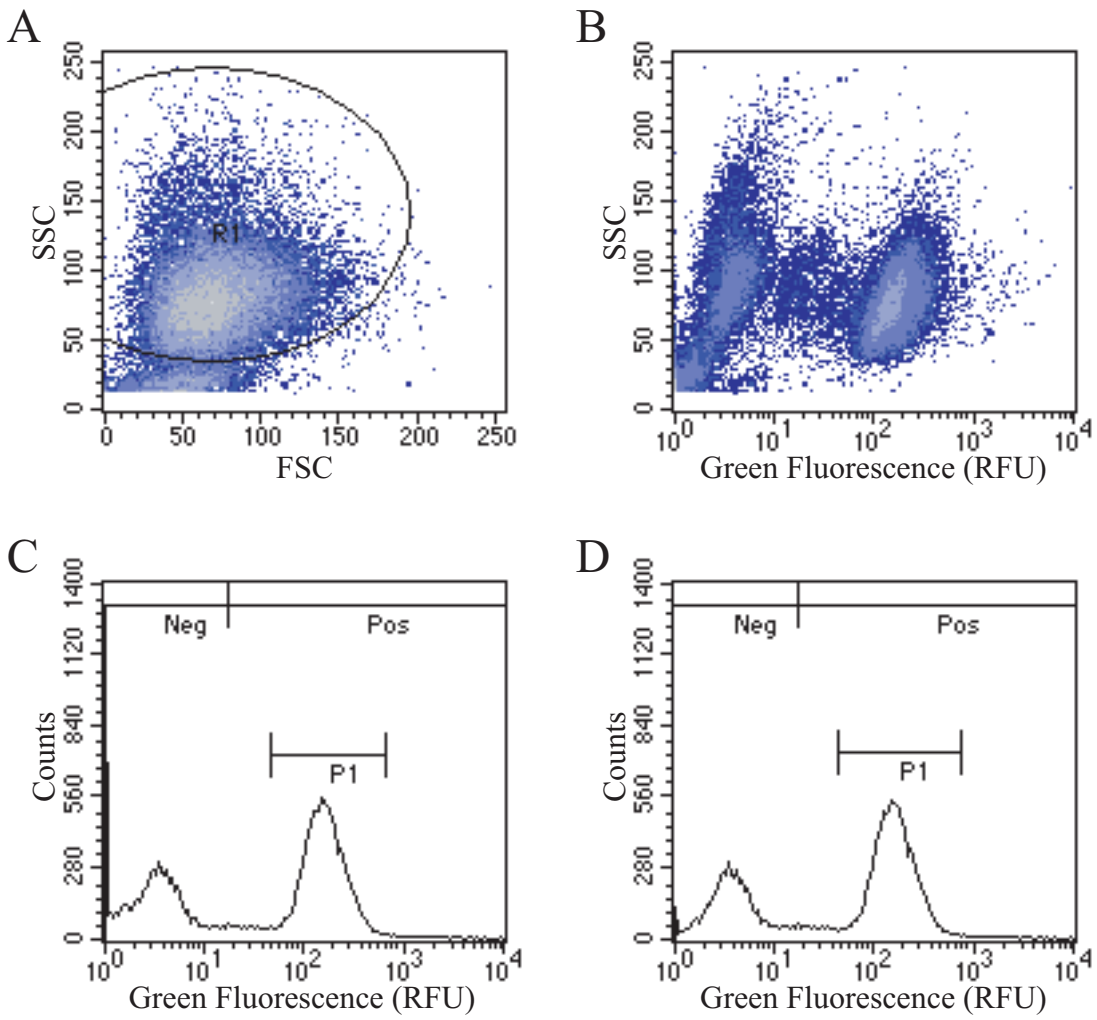


Figure 2-5. Green Fluorescence and Light Scatter Properties of Transgenic NT-1 Protoplast Preparations.
(A) Biparametric flow analysis of transgenic protoplasts, based on side light scatter and forward light scatter. Gated events (R1) are plotted in (D). **(B)** Biparametric flow analysis of transgenic protoplasts, based on side scatter and green fluorescence. **(C)** Green fluorescence histogram of all data from transgenic protoplasts in (A). **(D)** Green fluorescence histogram of data from gated transgenic protoplasts (R1) from (A). Parameters for histograms are the same as in figure 2-3.

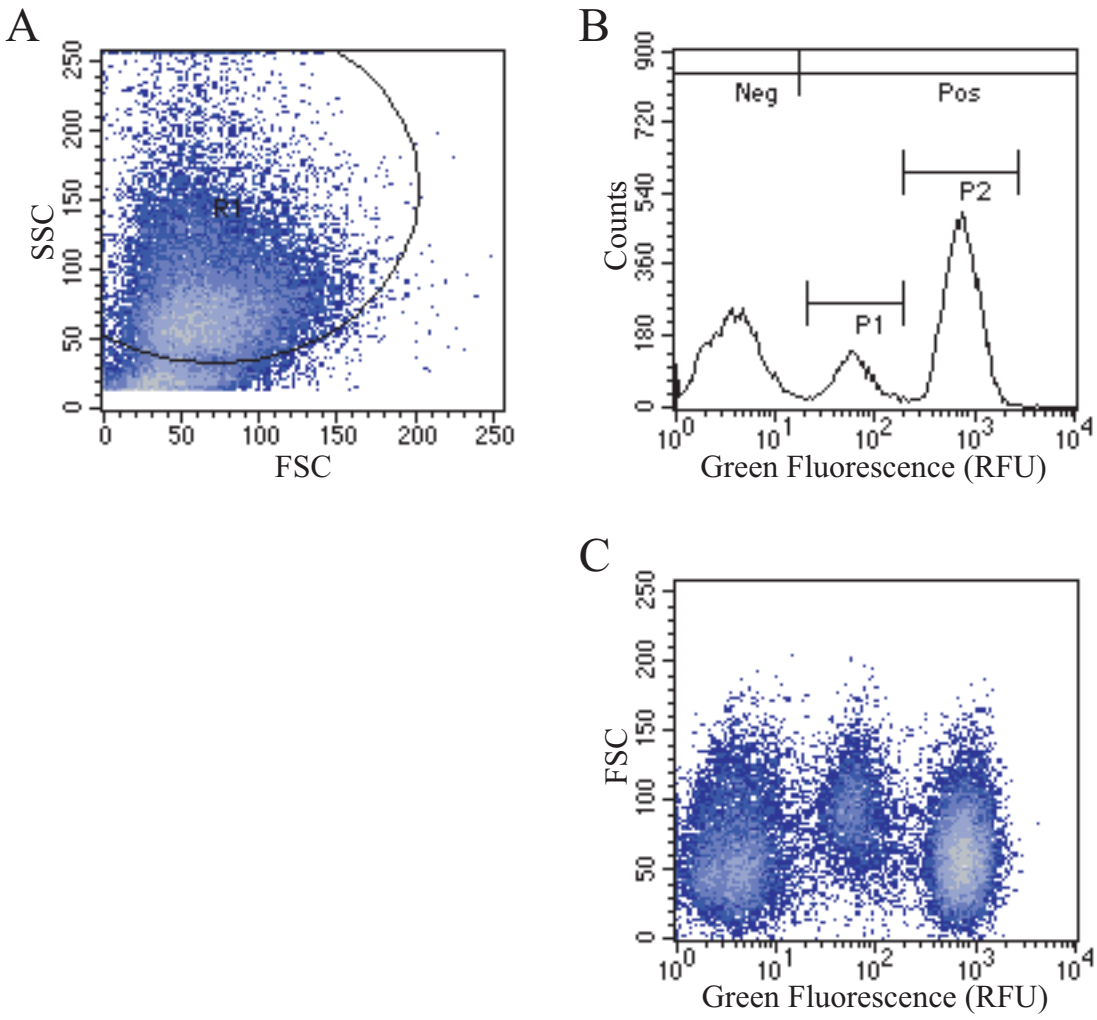


Figure 2-6. Damaged Protoplasts May Contribute to an Underestimation of GFP Expression.
(A) Biparametric flow analysis of transgenic protoplasts, based on side light scatter and forward light scatter. Gated events (R1) are plotted in (B). **(B)** Green fluorescence histogram of gated transgenic protoplasts (R1) from (A). **(C)** Biparametric flow analysis of transgenic protoplasts, based on forward scatter and green fluorescence demonstrating uncharacteristic increased forward scatter in protoplasts composing P1 in (B). Parameters for histograms are the same as in figure 2-3.

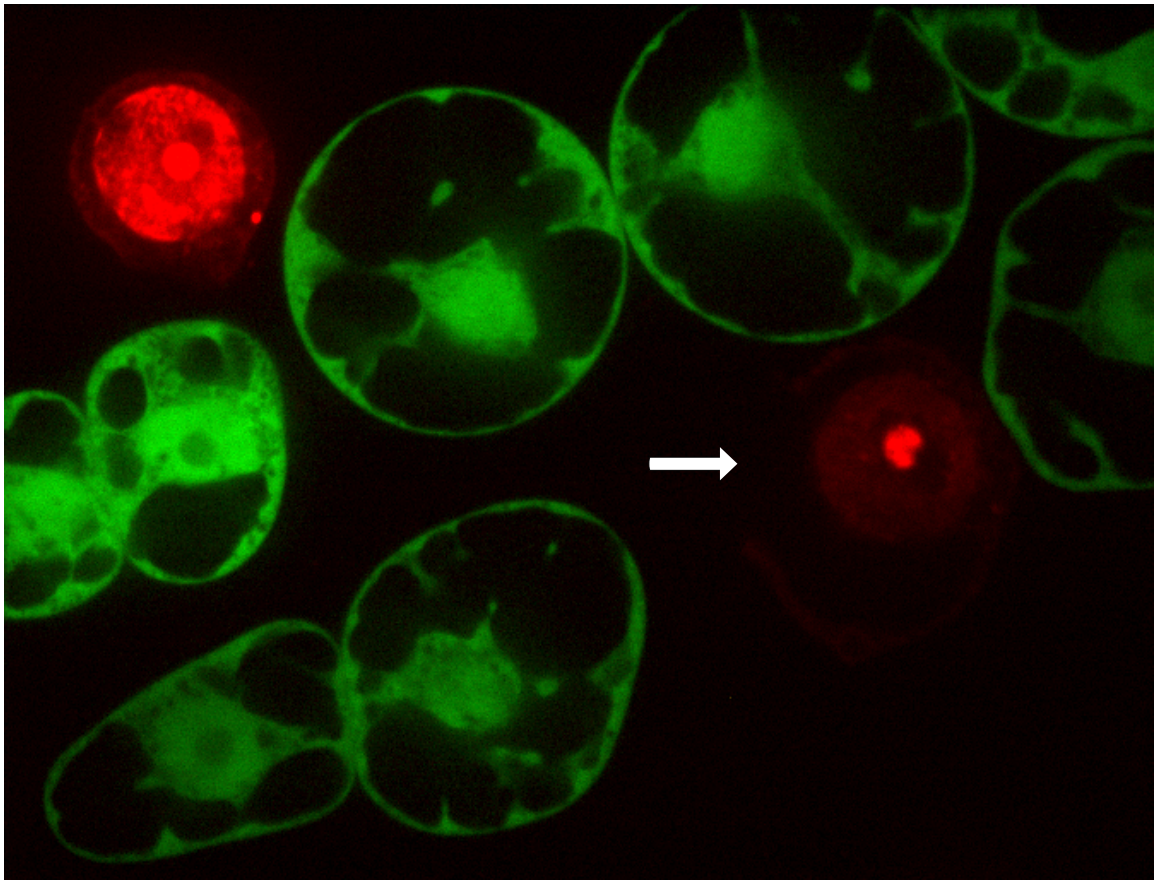


Figure 2-7. Propidium Iodide Stains Free Nuclei and Nuclei in Protoplasts with Compromised Plasma Membrane Integrity.
A 1 μm thick confocal section of a propidium iodide-stained protoplast preparation of a tobacco cell line uniformly expressing GFP. Intact protoplasts exclude propidium iodide while free nuclei and nuclei in protoplasts with compromised plasma membrane integrity stain brightly. GFP is absent from protoplasts in which nuclei stain brightly with propidium iodide. A white arrow indicates a missing portion of the plasma membrane.

Figure 2-8. Propidium Iodide Staining of Protoplast Preparations Improves Flow Cytometry Resolution.
(A) Transgenic protoplasts stained with PI subjected to biparametric analysis of side light scatter versus forward light scatter. R1 gated events are plotted in (B). **(B)** Green fluorescence histogram of gated protoplasts (R1) from (A). **(C)** The same data from the same transgenic protoplasts stained with PI were subjected to biparametric analysis of side light scatter versus PI relative orange fluorescence. R2 gated events are plotted in (D). **(D)** Green fluorescence histogram of gated protoplasts (R2) from (C). Note: after exclusion of protoplasts and debris that have high levels of PI fluorescence, the mode consistent with wild-type fluorescence is lost. **(E)** A separate protoplast preparation from the same transgenic cell line subjected to biparametric analysis of side light scatter versus forward light scatter. R3 gated events are plotted in (F). **(F)** Green fluorescence histogram of gated protoplasts (R3) from (E). Note: gating based on light scatter properties alone is sufficient to exclude the mode consistent with wild-type fluorescence in this example. Parameters for histograms are the same as in figure 2-3.

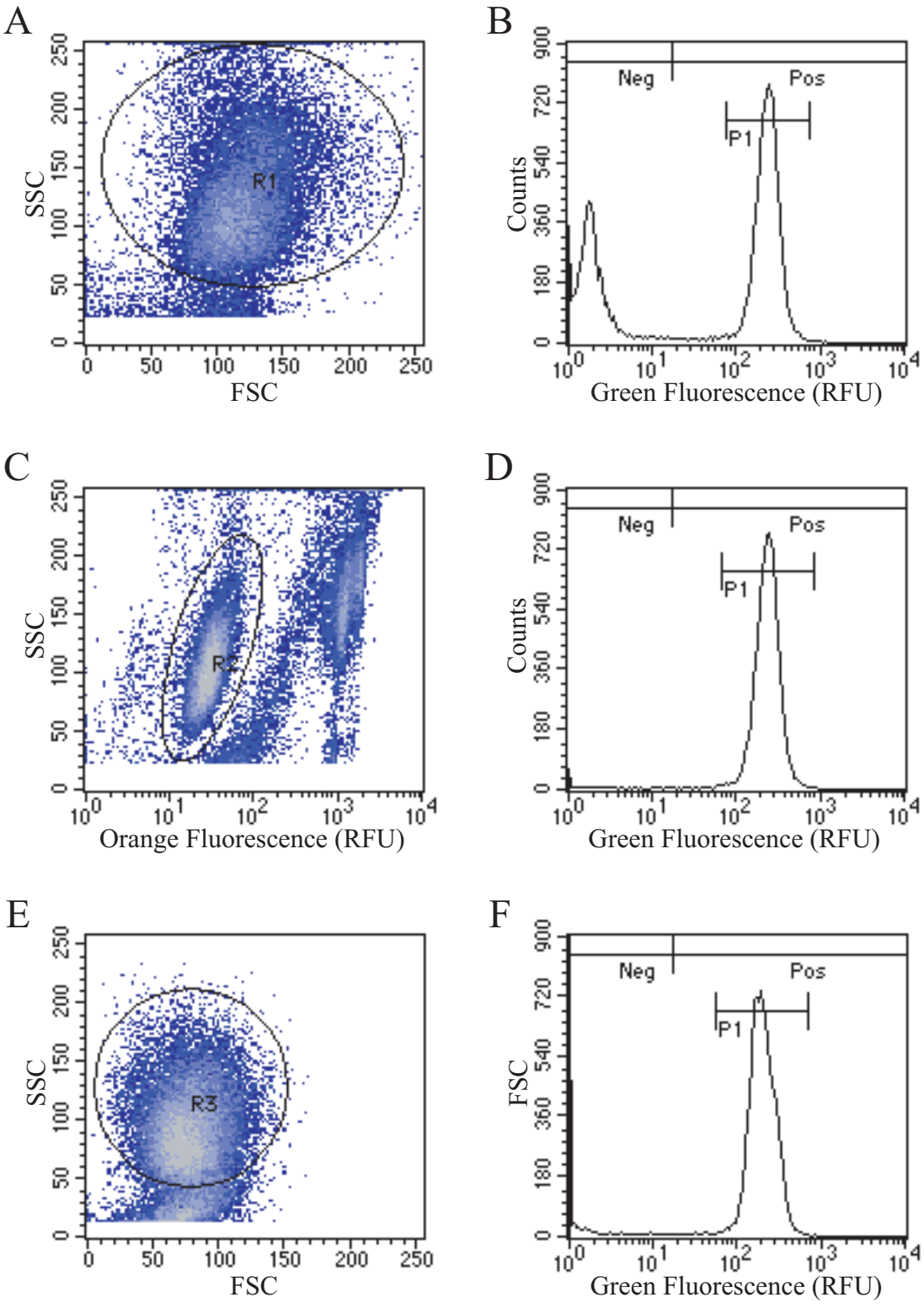

A
SSC
FSC
R1
B
Counts
Green Fluorescence (RFU)
Neg
Pos
P1
C
SSC
Orange Fluorescence (RFU)
R2
D
Counts
Green Fluorescence (RFU)
Neg
Pos
P1
E
SSC
FSC
R3
F
FSC
Green Fluorescence (RFU)
Neg
Pos
P1

Figure 2-9. Propidium Iodide Stains Protoplasts with Uncharacteristically High Levels of Forward Light Scatter.
(A) Biparametric flow analysis of transgenic protoplasts, based on side light scatter and forward light scatter. Gated events (R1) are plotted in (B). **(B)** Green fluorescence histogram of gated transgenic protoplasts (R1) from (A). **(C)** Biparametric flow analysis of transgenic protoplasts, based on forward scatter and green fluorescence demonstrating uncharacteristic increased forward scatter in protoplasts composing P1 in (B). **(D)** Biparametric flow analysis of the same transgenic protoplasts stained with PI comparing green and orange fluorescence. Protoplasts with intermediate levels of green fluorescence (P1 in B) have high levels of orange PI fluorescence similar to debris. **(E)** Transgenic protoplasts stained with PI subjected to biparametric analysis of side light scatter versus PI relative orange fluorescence. R2 gated events are plotted in (F). **(F)** Green fluorescence histogram of gated protoplasts (R2) from (E). Parameters for histograms are the same as in figure 2-3.

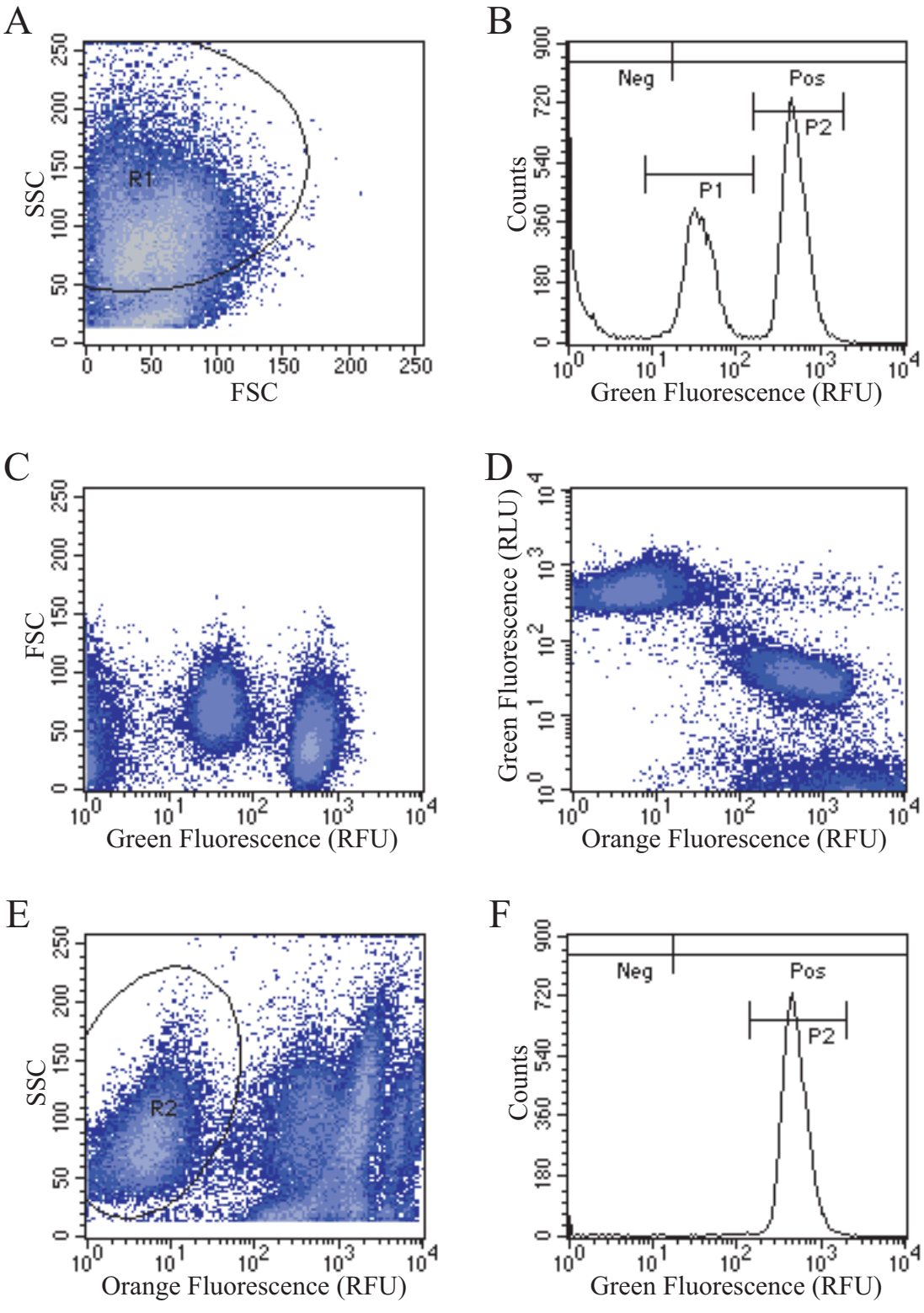

A
SSC
FSC
R1
B
Counts
Green Fluorescence (RFU)
Neg
Pos
P1
P2
C
FSC
Green Fluorescence (RFU)
D
Green Fluorescence (RLU)
Orange Fluorescence (RFU)
E
SSC
Orange Fluorescence (RFU)
R2
F
Counts
Green Fluorescence (RFU)
Neg
Pos
P2

Figure 2-10. Propidium Iodide Concentration Comparison.
(A) Transgenic protoplasts stained with a final concentration of 5 µg/mL PI were subjected to biparametric analysis of side light scatter versus PI relative orange fluorescence. R1 gated events are plotted in (B). **(B)** Green fluorescence histogram of gated protoplasts (R1) from (A). **(C)** and **(D)** Same analysis as (A) and (B) except protoplasts were stained with a final concentration of 2.5 µg/mL or similarly stained with 1.25 µg/mL **(E)** and **(F)**. Parameters for histograms are the same as in figure 2-3.

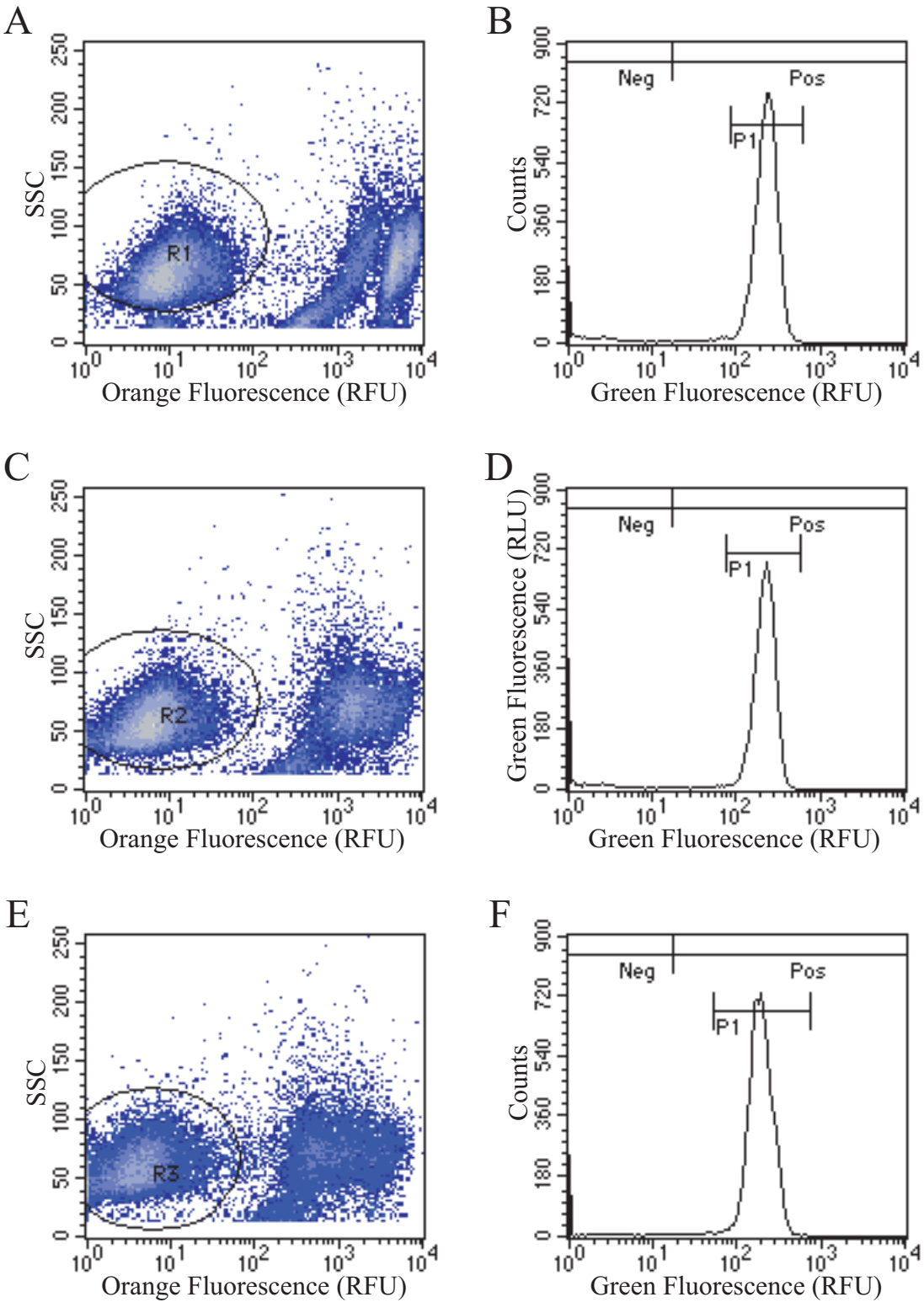
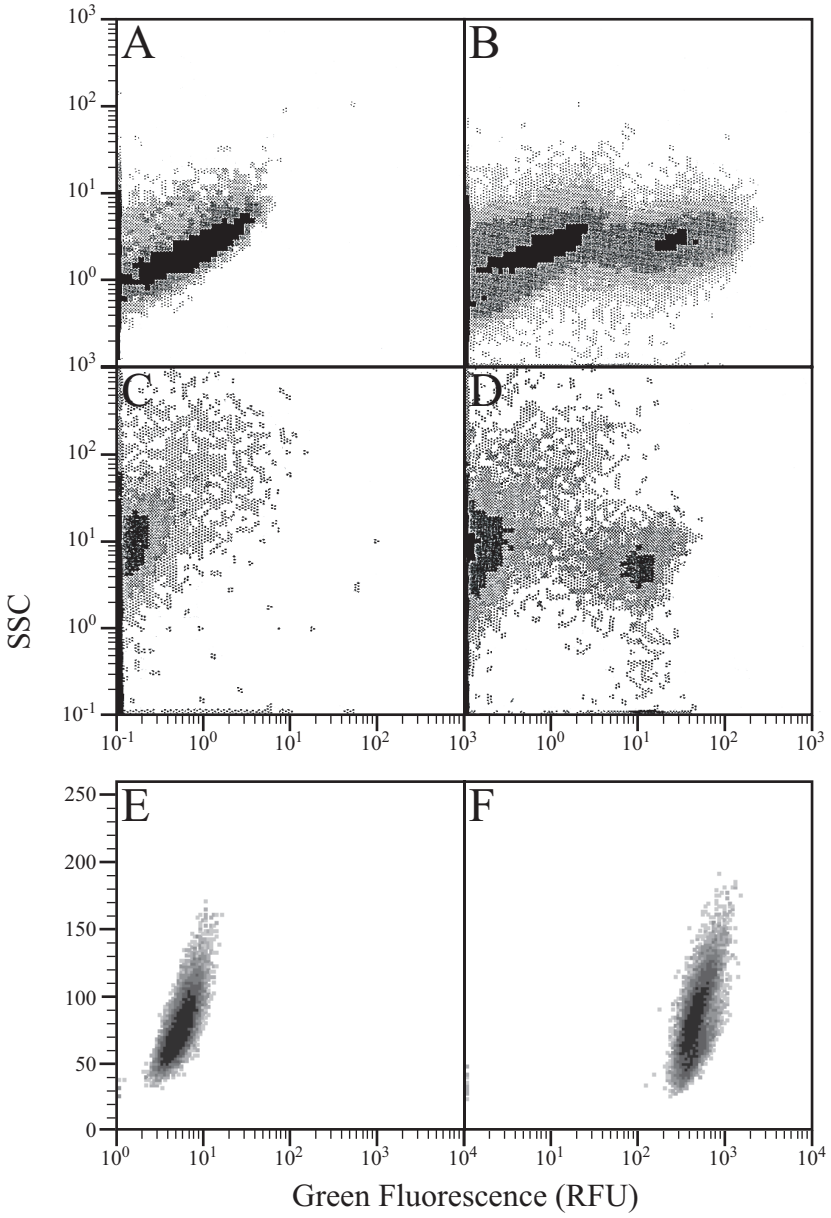

A
SSC
R1
Orange Fluorescence (RFU)
B
Neg
Pos
P1
Counts
Green Fluorescence (RFU)
C
SSC
R2
Orange Fluorescence (RFU)
D
Neg
Pos
P1
Green Fluorescence (RLU)
Green Fluorescence (RFU)
E
SSC
R3
Orange Fluorescence (RFU)
F
Neg
Pos
P1
Counts
Green Fluorescence (RFU)

Figure 2-11. Comparison with Current Flow Cytometric Methods of Detection of Green Fluorescence.
(A) Biparametric flow analysis of untransfected (control) etiolated maize protoplasts, based on side light scatter and green fluorescence. **(B)** Biparametric flow analysis of etiolated maize protoplasts transfected with a GFP construct, based on side light scatter and green fluorescence. (Data from A and B from Galbraith et al., 1995) **(C)** Biparametric flow analysis of tobacco nuclei isolated from untransformed plants, based on side light scatter and green fluorescence. **(D)** Biparametric flow analysis of tobacco nuclei isolated from transgenic plants transformed with a GFP:GUS fusion construct, based on side light scatter and green fluorescence. Note: the added size of the GFP:GUS fusion prevents passive diffusion through nuclear pores. (Data from C and D from Grebenok et al., 1997) **(E)** Biparametric flow analysis of untransformed NT-1 tobacco protoplasts from propidium iodide gating methods presented in this work, based on side light scatter and green fluorescence. **(F)** Biparametric flow analysis of transgenic NT-1 tobacco protoplasts transformed with pGFP from propidium iodide gating methods presented in this work, based on side light scatter and green fluorescence.


A
B
C
D
E
F
SSC
Green Fluorescence (RFU)

References

An, G.H. (1985). High-efficiency transformation of cultured tobacco cells. Plant Physiol **79,** 568-570.

Becker, D., Kemper, E., Schell, J., and Masterson, R. (1992). New plant binary vectors with selectable markers located proximal to the left T-DNA border. Plant Mol Biol **20,** 1195-1197.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science **302,** 1956-1960.

Bohanec, B., Luthar, Z., and Rudolf, K. (2002). A protocol for quantitative analysis of green fluorescent protein-transformed plants, using multiparameter flow cytometry with cluster analysis. Acta Biologica Cracoviensia Series Botanica **44,** 145-153.

Callaway, A., Liu, W., Andrianov, V., Stenzler, L., Zhao, J., Wettlaufer, S., Jayakumar, P., and Howell, S.H. (1996). Characterization of cauliflower mosaic virus (CaMV) resistance in virus-resistant ecotypes of Arabidopsis. Mol Plant Microbe Interact **9,** 810-818.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science **263,** 802-805.

Crameri, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol **14,** 315-319.

Davis, S.J., and Vierstra, R.D. (1998). Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol Biol **36,** 521-528.

Desikan, R., Hagenbeek, D., Neill, S.J., and Rock, C.D. (1999). Flow cytometry and surface plasmon resonance analyses demonstrate that the monoclonal antibody JIM19 interacts with a rice cell surface component involved in abscisic acid signalling in protoplasts. FEBS Lett **456,** 257-262.

Doran, P.M. (2000). Foreign protein production in plant tissue cultures. Curr Opin Biotechnol **11,** 199-204.

Galbraith, D.W., Herzenberg, L.A., and Anderson, M.T. (1999a). Flow cytometric analysis of transgene expression in higher plants: green fluorescent protein. Methods Enzymol **302,** 296-315.

Galbraith, D.W., Anderson, M.T., and Herzenberg, L.A. (1999b). Flow cytometric analysis and FACS sorting of cells based on GFP accumulation. Methods Cell Biol **58,** 315-341.

Galbraith, D.W., Lambert, G.M., Grebenok, R.J., and Sheen, J. (1995). Flow cytometric analysis of transgene expression in higher plants: green-fluorescent protein. Methods Cell Biol **50,** 3-14.

Grebenok, R.J., Lambert, G.M., and Galbraith, D.W. (1997). Characterization of the targeted nuclear accumulation of GFP within the cells of transgenic plants. Plant J **12,** 685-696.

Haseloff, J., Siemering, K.R., Prasher, D.C., and Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. Proc Natl Acad Sci U S A **94,** 2122-2127.

Heim, R., Prasher, D.C., and Tsien, R.Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci U S A **91,** 12501-12504.

Heim, R., Cubitt, A.B., and Tsien, R.Y. (1995). Improved green fluorescence. Nature **373,** 663-664.

Jaroszeski, M.J., and Heller, R. (1998). Flow cytometry protocols. (Totowa, N.J.: Humana Press).

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

Mankin, S.L. (2000). Chromatin elements and transgene expression in tobacco cells [dissertation]. In Botany (Raleigh, NC: North Carolina State University).

McGann, L.E., Walterson, M.L., and Hogg, L.M. (1988). Light scattering and cell volumes in osmotically stressed and frozen-thawed cells. Cytometry **9,** 33-38.

Migliaccio, A.R., Bengra, C., Ling, J., Pi, W., Li, C., Zeng, S., Keskintepe, M., Whitney, B., Sanchez, M., Migliaccio, G., and Tuan, D. (2000). Stable and unstable transgene integration sites in the human genome: extinction of the green fluorescent protein transgene in K562 cells. Gene **256,** 197-214.

Piston, D.W., Patterson, G.H., and Knobel, S.M. (1999). Quantitative imaging of the green fluorescent protein (GFP). Methods Cell Biol **58,** 31-48.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. Gene **111,** 229-233.

Shapiro, H.M. (1995). Practical flow cytometry. (New York: Wiley-Liss).

Sieburth, L.E., and Meyerowitz, E.M. (1997). Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. Plant Cell **9,** 355-365.

CHAPTER 3

The Rb7 Matrix Attachment Region Increases the Likelihood and Magnitude of Transgene Expression

Introduction

In recent years, gene transfer into a wide variety of plant species has become almost common practice, but lack of predictable and reliable transgene expression has plagued efforts to fully exploit gene transfer technology for plant improvement and basic research. Initially, major differences in transgene expression between independent primary transformants were attributed to transgene silencing caused by integration of transgenes into or near heterochromatin regions, a phenomenon similar to "position effects" that have been well characterized in *Drosophila* (reviewed by Weiler and Wakimoto, 1995). We now know that gene silencing is more complex, and although we are still far from understanding variability in the silencing of transgenes, great progress has been made in understanding transcriptional gene silencing (TGS), post transcriptional gene silencing (PTGS) and the interplay between the two (Carrington and Ambros, 2003; Grewal and Moazed, 2003; Matzke and Matzke, 2003; Schramke and Allshire, 2003).

It is clear that in order to derive maximum benefits from the use of transgenic plants for biotechnology and basic research, gene silencing must be understood and methods to combat gene silencing must be devised. One approach that has shown promise in combating gene silencing involves the use of Matrix Attachment Regions (MARs) to flank transgenes (Allen et al., 2000). MARs are AT-rich DNA sequences that

bind to the nuclear matrix *in vitro* and are thought to organize chromosomal DNA into loop domains (Bonifer et al., 1990; Bode et al., 1996; Martelli et al., 2002). Early experiments in animal cells indicated that MARs increased transgene expression and decreased transformant-to-transformant variability in transgene expression (Stief et al., 1989; Bonifer et al., 1990; Phi-Van et al., 1990). Subsequent work in plant systems also demonstrated an increase in transgene expression and/or a reduction in variability (Breyne et al., 1992; Allen et al., 1993, 1996; Mlynarova et al., 1994, 1995; Han et al., 1997; Liu and Tabe, 1998; Ülker et al., 1999; Vain et al., 1999; Cheng et al., 2001; Mendu et al., 2001; Mankin et al., 2003). The observations of increases in expression and/or reductions in variability have not, however, been consistent. Some reports indicate that MARs have no effect on transgene expression (van Leeuwen et al., 2001; De Bolle et al., 2003; Sidorenko et al., 2003), and in at least two cases, a decrease in transgene expression has been reported (Breyne et al., 1992; Holmes-Davis and Comai, 2002).

Direct comparison of the results of the different studies are difficult to make, because various plant species, MARs, transformation methods, promoters and reporter genes have been used. Both cultured cells and intact plants have served as experimental material, and a wide variety of tissues in various stages of development have been sampled. In nearly all cases, histochemical approaches and/or extractive approaches have been used to assess activity of reporter genes. Histochemical approaches are not quantitative, but they do reveal cell-to-cell epigenetic variation in reporter gene expression. Approaches in which the product of the reporter gene is extracted from

tissues and then measured are quantitative but they obscure epigenetic cell-to-cell differences. Attempts have been made using the firefly luciferase reporter gene to simultaneously quantitate transgene expression and spatial variation in transgene expression in whole tobacco plants (van Leeuwen et al., 2001). In these studies, however, transgene expression was not measured at the resolution of individual cells.

In histochemical assays for transgene expression in both cultured cells and in whole plants, we have noticed that the expression is often variegated (Allen et al., 1993; Ülker et al., 1999), i.e., some cells appear to express the transgene and others do not. This raises the possibility that the MAR-mediated elevation in transgene expression measured by extractive procedures is due to epigenetic changes in the distribution of expressing cells within a population. That is, increases in transgene expression may reflect a greater proportion of cells expressing the transgene rather than an increase in the level of expression in expressing cells.

To test this possibility, we have used flow cytometry to measure green fluorescent protein (GFP) expression in individual cultured tobacco cells from clonal lineages of cells transformed by *Agrobacterium tumefaciens*. Binary vectors used for transformation carried the GFP reporter gene flanked on both sides by the tobacco Rb7 MAR, which we have previously used (Allen et al., 1996; Ülker et al., 1999; Mendu et al., 2001; Mankin et al., 2003). GFP levels were measured in protoplasts generated from the various transgenic lines and compared to levels in protoplasts generated from lines carrying control constructs without MARs. We discovered that the cell lines carrying constructs including the Rb7 MAR exhibited both a higher percentage of cells expressing GFP and a

higher level of GFP expression in the expressing cells. These effects were observed not only by compiling the results from many individual cell lines but also by making measurements on large pools of transformed cell lines. Our results suggest that the Rb7 MAR acts both to reduce the proportion of non-expressing cells and to increase the levels of expression in expressing cells.

Methods

Recombinant Plasmids

Standard cloning techniques (Sambrook et al., 1989) were used to create three binary vectors with T-DNA regions containing GFP described in figure 3-1. The first control construct (pGFP) was kindly provided by Luke Mankin and originally described as pLMNC-b47 (Mankin, 2000). This binary vector (pGFP) contains the GFP cassette from psmRS-GFP (Davis and Vierstra, 1998) as well as the *nptII* kanamycin resistance gene cassette and binary vector backbone from pGPTV-KAN (Becker et al., 1992). In order to generate an additional control construct (pLGFPL), the GFP cassette from psmRS-GFP (Davis and Vierstra, 1998) was cloned into pLMNC-b73 (Mankin, 2000) replacing the GUS cassette and resulting in GFP flanked by direct repeats of a 1.2 kb region of Phage Lambda DNA (GenBank accession number J02459, nucleotides 21231-22425). Similarly the MAR containing construct (pMGFPM) was made by cloning the same GFP cassette into pLMNC-b72 (Mankin, 2000) replacing the GUS cassette and resulting in GFP flanked by direct repeats of the 1.2 kb *Nicotiana tabaccum* Rb7 MAR

fragment from pRb7-6 (Hall et al., 1991). All T-DNA constructs contained the same orientation of gene cassettes and the same vector backbone sequences.

Plant Cell Transformation and Propagation

T-DNA vectors were electroporated into an *Agrobacterium tumefaciens* strain (UIA143/ pMP90) described by Callaway et al. (1996). To examine the effect of the Rb7 MAR on GFP expression in stable transformants, NT-1 cells from tobacco suspension cultures were maintained and transformed as described (An, 1985; Allen et al., 1993). These cells contain essentially no chlorophyll. Transformants were selected on solid NT-1 medium (An, 1985) supplemented with 200 mg/L Timintin (*Phyto*Technology Laboratories, Shawnee Mission, KS) and 50 mg/L kanamycin (Sigma, St. Louis, MO). Microcalli were clearly visible after two weeks, and 72 putative transformants per construct were isolated by transferring the microcalli to a second plate of fresh medium. After two weeks, a small, randomly selected, piece of the resulting callus tissue (approximately 1 mm in diameter) was transferred to a final third plate of fresh medium. After two additional weeks, the entire callus of 30 randomly selected calli was dispersed in 5 mL of fresh liquid medium supplemented with 200 mg/L Timintin and 50 mg/L kanamycin. Liquid cultures were grown at 27ºC under constant light conditions on an orbital shaker (125 rpm) and transferred weekly by adding 0.5 mL to 5 mL of fresh medium.

During the growth on the second fresh medium plate previously mentioned, many calli exhibited a sectored GFP phenotype. In order to investigate the possibility that the

sectored calli were chimeric (the result of multiple transformation events), calli of a subset of these were further investigated. While under far blue light (described below), small pieces of calli (approximately 1 mm in diameter) representing different sectors were transferred to a third plate of fresh medium. After three weeks, liquid cultures were initiated as mentioned previously. Southern blots of DNA from cultures arising from the different sectors were used to determine if sectors arose from separate transformation events.

Population based liquid cultures were initiated by resuspending the original Agrobacterium and NT-1 co-cultivation plate with 4 mL of media supplemented with 200 mg/L Timintin and 50 mg/L kanamycin.

GFP Imaging

Confocal sections of protoplasts were obtained with a Perkin Elmer Ultra*VIEW* live cell imager (Perkin Elmer, Boston, MA). NT-1 calli were imaged every 4 days throughout the course of solid medium selection. A Nikon Coolpix 990 (370 nm filter) was attached to a tripod and positioned vertically over petri plates containing NT-1 callus. Two Clare Chemical Research Dark Reader™ (Model DL-09B) hand lamps (Clare Chemical Research, Inc., Dolores, CO) were used to excite GFP fluorophors with far blue light (~420 nm to 500 nm). Images of the calli were processed in Adobe Photoshop to remove signal from red and blue channels.

Sample Preparation, Flow Cytometry, and Analysis

Existing techniques in flow cytometry and protoplast preparation methods (Galbraith et al., 1995, 1999a, b) were adapted and modified for our specific application. Protoplasts were obtained from four day old NT-1 liquid cultures by the following procedure. The protoplasting solution was composed of 0.4 M manitol, 10 mM MES pH 5.7, 1 mg/mL Pectolyase Y-23 and 10 mg/mL Cellulase RS (Karlan Research Products Corporation, Santa Rosa, CA). 300 µL of pelleted cells were suspended in a 0.4 M manitol, 10 mM MES pH 5.7 solution, pelleted at 450 g for 3 minutes, and resuspended in 500 µL of protoplasting solution. Incubation for 45 minutes at 28ºC on a rotating shaker (125 rpm) yielded >75% liberated protoplasts. The remaining steps were carried out at 4ºC. Protoplasts were centrifuged for 3 minutes at 100 g, washed in a 0.4 M manitol, 10 mM MES pH 5.7 solution, and resuspended in 1 mL of a 0.4 M manitol, 10 mM MES solution. Samples were then allowed to pass through a 50 µm mesh and stained with 5 µL of 1 mg/mL propidium iodide prior to flow cytometric analysis.

Protoplasts were stored on ice a maximum of two hours until flow cytometry was conducted. The flow cytometer FACSCalibur (Becton-Dickinson, San Jose, CA) was operated at a laser wavelength of 488 nm with a power output of 15 mW, using a sort-sense flow cell equipped with a 100 nm orifice. Forward angle light scatter (FSC) signals, side angle light scatter (SSC) signals, and fluorescence emission signals corresponding to wavelengths of 520-530 nm (green) and 564-606 nm (yellow/orange) were collected at a maximum flow rate of 1,000 cells/sec with a total accumulation of 15,000 viable cells based on populations of minimal propidium iodide staining. In order

to collect data on 15,000 viable cells, approximately 30,000 total events were needed depending on the effectiveness of the protoplasting procedure for each sample. The FSC detector was operated at a gain setting of 9.9 and a high voltage of E-1V. The SSC detector high voltage was set to 319 V with no gain adjustments. The PMT high voltages were 370-450 V (green channel) based on calibration beads (discussed below) and 447 V (yellow/orange channel) with no gain adjustments for either. Uniparametric and biparametric displays were generated with CellQuest 3.3 (Becton-Dickinson, San Jose, CA).

To insure that degradation of laser intensity and other instrument-specific fluctuations did not affect our results taken over several time points, Flow Check™ YG 6.0 microspheres (Polysciences Inc., Warrington, PA) were used to calibrate the flow cytometer before each use. Three distinct intensities of green fluorescence provided three calibration points to adjust the PMT voltage of the green channel. Because the fluorescence intensities of the microspheres in one set may not match those of a second set from the same manufacturer, the PMT voltage of the green channel was adjusted to match calibration points for the original set, and then with the same instrument settings, microspheres from the second set were analyzed to determine the new calibration points for this second set.

Since neither green fluorescence intensity data nor percentage of GFP positive cells were normally distributed as determined by the Kolmogorov-Smirnov test (Chakravarti et al., 1967), a non-parametric Mann-Whitney test was used to compare treatments and calculate p-values (Snedecor and Cochran, 1989).

DNA Extraction and Southern Blot Analysis

DNA was isolated from frozen tissue collected from transgenic tobacco cells using a modified CTAB extraction procedure (Murray and Thompson, 1980) as described by Johnson et al. (1995).

Fifteen µg of genomic DNA was digested with *Hin*dIII (Promega, Madison, WI), a methylation insensitive restriction enzyme, for 12 hr at 37ºC in a reaction volume of 50 µL containing 30 units of restriction enzyme according to the instructions of the manufacturer. The fragments were then separated by electrophoresis for 16 hr in a 1% agarose gel in TAE and blotted onto MagnaGraph nylon membranes (Osmonics Inc., Westborough, MA) by overnight capillary transfer. Membranes were pre-hybridized overnight at 65ºC with hybridization buffer from an AlkPhos Direct kit (Amersham Biosciences Corp., Piscataway, NJ) and hybridized overnight in the same buffer at 65ºC with a 458 bp PCR amplified fragment immediately internal to the right border of all T-DNAs tested. The primers used were T-DNA RB probe FWD (5'-CCAATTCGCCCTATAGTGAGTCGT-3') and T-DNA RB probe REV (5'-CATGCACATACAAATGGACGAACGG-3'). After PCR amplification the probe was ^{32}P-labelled by random-primed synthesis with Rediprime II (Amersham Biosciences Corp., Piscataway, NJ). Membranes were washed five times with the final two washes being the most stringent consisting of 0.5X SSC and 0.1% SDS at 65ºC for 20 minutes each. Hybridized membranes were then placed in phosphor screens for varying amounts of time and read on a Storm 840 phosphoimager (Amersham Biosciences Corp.,

Piscataway, NJ). Similarly, membranes were probed with 256 bp PCR amplified fragment from the 18S rRNA coding region of NT-1 genomic DNA with primers 18S rDNA FWD (5'-ACCAGACTCATAGAGCCCGGTATT-3') and 18S rDNA REV (5'-TCATGATAACTCGACGGATCGCAC-3').

Results

In order to test the effect of MARs on green fluorescent protein (GFP) expression in *Agrobacterium*-transformed tobacco cells in culture, we used a T-DNA construct in which the GFP cassette was flanked on both sides by the Rb7 MAR. This construct is denoted "MGFPM" (Figure 3-1). We compared results for this construct with results for two control constructs. In the first control ("GFP"), the GFP cassette is inserted into the T-DNA alone. In the second control, the GFP cassette is flanked by a 1.2 kb region of phage lambda DNA that does not bind to tobacco nuclear matrices (Mendu and Spiker, unpublished results). This control ("LGFPL") allows us to discriminate sequence-dependent MAR effects from effects that may be due to MARs acting simply as spacer elements to distance the GFP cassette from neighboring T-DNA and host genomic DNA. The soluble-modified, red-shifted GFP with its enhanced fluorescence (Davis and Vierstra, 1998) was used, since it was found to be more suitable for flow cytrometry (Galbraith et al., 1999b).

Improved Flow Cytometry Resolution by Debris Discrimination

Because tobacco suspension cells grow in filamentous structures that cannot be analyzed by flow cytometry, it is necessary to disrupt the cell walls that hold these filaments together. A gentle digestion with cellulases and pectolyases releases individual cells as protoplasts that can pass one at a time through the flow cell of the cytometer to measure GFP expression in individual cells. The protoplasting procedure creates some debris, such as free nuclei and some protoplasts with compromised plasma membrane integrity. If these entities were included in our analyses, spurious results would be obtained, because GFP diffuses out of free nuclei and compromised protoplasts. GFP-expressing protoplasts with damaged membranes would be counted as non-expressing protoplasts, leading to an underestimate of GFP in the cell population.

In order to illustrate these phenomena, a cell line was identified that had high levels of uniform, non-variegated, GFP fluorescence prior to protoplasting. As determined by microscopy, all cells in this cell line had levels of GFP fluorescence that were indistinguishable from one another (data not shown). Protoplasts were prepared from this cell line, stained with propidium iodide, and subjected to confocal microscopy (Figure 3-2A). Clearly, GFP fluorescence is no longer uniform for all protoplasts. Protoplasts with compromised plasma membrane integrity allow GFP to diffuse out of the cell and propidium iodide to diffuse in and stain the nucleus. Despite having defects in plasma membrane integrity, these protoplasts may still be able to maintain their size and general shape. Spurious flow cytometry data from these protoplasts and from free

nuclei can be excluded from analysis by measuring and accounting for the high levels of orange fluorescence associated with propidium iodide staining.

Intrinsic physical parameters such as forward light scatter, a measure of particle size, and side light scatter, a measure of cytoplasmic granularity (Shapiro, 1995), can be used to discriminate between clusters of light scatter values from cells and debris such as damaged cells that often have reduced light scatter properties (McGann et al., 1988). Some of the data from debris in our protoplast preparations can thus be excluded by gating, i.e., including only data points consistent with the light scattering properties of protoplasts, as defined by an oval outline in figure 3-2C. When gating based only on light scattering is used, it appears as though a substantial proportion of the cells (32% in this example) have low levels of fluorescence equivalent to those of wild type cells (compare figure 3-2D with figure 3-2B). Because microscopic evaluation of the cells prior to protoplasting showed that all cells displayed uniform GFP fluorescence, it appears that the protoplasts with compromised membranes and residual nuclei not gated away cause an overestimation of the proportion of cells that have no fluorescence (Figure 3-2D). This problem can be alleviated by gating data based on both side light scatter (to exclude data from debris that is not consistent with light scatter properties of protoplasts) and from orange fluorescence associated with propidium iodide staining (to exclude data from free nuclei and protoplasts with compromised membranes that have allowed GFP to diffuse out of the cell and propidium iodide to diffuse in and stain nuclei). When the exact same data as shown in figure 3-2C and 3-2D were gated based on side light scatter and orange fluorescence (Figure 3-2E), only 0.9% of the cells had fluorescence properties

consistent with wild type cells (Figure 3-2F). With this approach, the error associated with false-negative cells can be greatly reduced and the ability to quantitate the proportion of cells with no detectable GFP fluorescence greatly improved. This gating approach was used in all analyses that follow.

Rb7 MAR Increases the Likelihood and Magnitude of GFP Expression in Individual Cell Lines

Three independent replicates of *Agrobacterium*-mediated transformation experiments were conducted to assess the effect of MARs on GFP expression in NT-1 cells in culture. Each replicate contained three T-DNA construct treatments to generate approximately 30 cell lines per T-DNA construct for a grand total of approximately 270 cell lines for all three replicates. The results in figure 3-3A represent the average GFP expression in an individual cell of all cell lines transformed with a given construct in each of the replicate experiments. The data used to generate this figure include all 15,000 protoplasts in each cell line that were present in a gated region (as in figure 3-2E). Cells that have green fluorescence properties characteristic of wild type cells are termed "GFP negative cells," while cells that have levels of green fluorescence greater than wild type cells are termed "GFP positive cells." Data from all cells, both GFP negative and GFP positive, are included in figure 3-3A. This type of analysis is analogous to transgene expression studies in which the product of the reporter gene is extracted from a large population of cells and measured quantitatively, ignoring cell to cell variation within cell lines. In the data shown in figure 3-3A, expression levels of the two controls (LGFPL

and GFP alone) are not significantly different ($p > 0.05$, see methods). Cell lines transformed with the MAR construct (MGFPM) had significantly higher GFP fluorescence ($p < 0.05$) when compared to lines carrying GFP alone. Fold increases of 1.99, 3.69 and 2.11 were observed in the three replicates.

The increase in GFP expression resulting from the use of the Rb7 MAR could be accounted for by two possible scenarios, which are not mutually exclusive: 1) a greater percentage of expressing cells in lines containing MGFPM (likelihood) and/or 2) higher levels of expression of individual cells in lines containing MGFPM (magnitude). In order to address the first possibility, the percentage of GFP positive cells was determined for each cell line. The averages of these percentages for all cell lines of each treatment in each of the three replicates are shown in figure 3-3B. The two control lines (LGFPL and GFP alone) are not statistically different ($p > 0.05$). Cell lines transformed with the MAR-containing construct (MGFPM) had higher percentages of positive cells than controls ($p < 0.05$) in replicate experiments one and two, with increases of 39% (from 55% to 94%) and 38% (from 59% to 97%) respectively compared to GFP alone. In replicate three, cell lines transformed with MGFPM had borderline significant ($p = 0.064$) increases in positive cells, with an increase of 27% (from 50% to 77%) compared to cell lines transformed with GFP alone. These results support the hypothesis that the presence of Rb7 MAR increases the likelihood of GFP expression.

In order to address the second possibility that the expression levels of individual expressing cells are higher in lines containing MGFPM, data were plotted for positive cells only. The averages of the magnitudes of GFP expression of expressing cells for cell

lines of each treatment in each of the three replicates are shown in figure 3-3C. The two control lines (LGFPL and GFP alone) are not statistically different ($p > 0.05$). In replicate two and three, cell lines transformed with MGFPM had significantly higher GFP expression than control lines ($p < 0.05$) with fold increases of 2.85 and 1.87 respectively. In replicate one, expressing cells in cell lines transformed with MGFPM had a higher average level of GFP expression than those transformed with control constructs, but the difference was not statistically significant ($p = 0.23$).

Assessment of the effect of MARs on the magnitude of GFP expression in the preceding paragraph was based on expressing cells in all cell lines. In lines with only a few expressing cells, it is possible that cells we have counted as positive could have arisen from noise associated with green fluorescence detection. To be conservative in our interpretation of the data, we have excluded lines with less than two percent GFP positive cells (13 GFP lines, 20 LGFP lines, and 4 MGFPM lines). With this conservative approach, expression of the two control constructs are still not statistically different ($p > 0.05$). In replicates two and three, cell lines transformed with MGFPM still had significantly higher GFP expression than control lines ($p < 0.05$) but had lower fold increases (2.73 and 1.74 respectively). In replicate one, expressing cells in cell lines transformed with MGFPM had a higher average level of GFP expression than those transformed with control constructs. The difference, however, was still not statistically significant and the p-value rose to 0.68. Even though our criteria for statistical significance was not met in all replicates, these results support the hypothesis that the

presence of the Rb7 MAR can increase the magnitude of GFP expression in expressing cells.

GFP Expression is Stable over Time with and without Continued Selection

In order to evaluate possible changes in GFP expression over time, cell lines were maintained with kanamycin selection after the first flow cytometry analysis was conducted. After four and eight additional months (with weekly passages), green fluorescence data were collected. No statistically significant changes in mean GFP expression were observed for any of the treatments (Figure 3-4). Similarly, no statistically significant changes in the percentages of GFP positive cells or the levels of GFP expression in positive cells were observed for any of the treatments (data not shown). Additionally, a subset of cell lines in each treatment were cultured without kanamycin selection over the final four months of the time course. Again, no changes in mean GFP expression levels or percentages of expressing cells were observed for any treatments (data not shown).

Rb7 MAR Increases the Likelihood and Magnitude of GFP Expression in Population Cell Cultures

In traditional approaches to determining the effect of specific DNA elements (e.g., promoters, enhancers, MARs), a number of separate cell lines are created and the averages of transgene expression in the lines are measured. This approach can be quite laborious. Since flow cytometry allows detection of GFP expression in individual cells,

it is possible to pool large populations of cells derived from many different transformation events and still retain an underlying distribution of GFP expression that would be lost if cell lines were pooled and assayed with a traditional extractive approach. This pooling strategy allows for the assessment of the effect of any specific DNA element on transgene expression across a wide range of *Agrobacterium*-mediated transformation events. Instead of isolating and propagating individual cell lines, entire *Agrobacterium* and tobacco cell culture co-cultivations were used to initiate liquid cultures under kanamycin selection. Cells from these pooled lines were subjected to the same flow cytometry analysis as outlined previously for individual cell lines. The results are shown in figure 3-5. Strikingly, MGFPM population culture lines have fewer GFP negative cells giving rise to an average of 92.3% GFP positive cells, while GFP and LGFPL lines collectively averaged only 66.1% expressing cells. This observation is very similar to the results in individual cell lines (Figure 3-3B). Another distinguishing feature is that population culture lines containing MGFPM have an additional intermediate mode of expression as well as an additional "shoulder" on the high end of the profile. The presence of this subset of cells expressing GFP at a higher level compared to control constructs is consistent with our observations in individual cell lines (Figure 3-3C).

Variegated GFP Expression Arises from Cells with the Same Genotype

In experiments involving the generation of individual cell lines that have been transformed with foreign DNA, it is desirable to use methods of DNA transformation and tissue propagation that eliminate or reduce the frequency of producing chimeric cell lines.

Even though care is taken to insure that each cell line arises from a single transformation event, it is possible that multiple transformation events could occur next to each other on culture plates and not be distinguished from a single event. If "individual" cell lines were composed of chimeric populations, conclusions about epigenetic variation in transgene expression could be compromised.

In our experiments, cell lines are originally identified as microcalli that appear as small "bumps" on a lawn of non-transformed tobacco NT-1 cells. It is likely that each microcallus results from a single transformation event as the microcalli are well separated on the culture plates. Nevertheless, the possibility that the microcalli are chimeric exists. The microcalli are picked and subcultured on solid media containing kanamycin before transfer to liquid media. During cell line propagation on solid media containing kanamycin, sectors of GFP expression were sometimes observed (Figure 3-6A, left column). Sectors were observed in all three rounds of solid selection (data not shown). Approximately one third of the calli displayed sectors of GFP fluorescence on the second plate of solid selection, whereas approximately one tenth of the calli displayed sectors of GFP fluorescence on the third plate of solid selection. In order to minimize the possibility of carrying chimeric lines into liquid culture, only small subsets of cells (pieces about 1 mm in diameter) are used to initiate new callus. Therefore our flow cytometry results are indicative of the variation that exists during the last solid state propagation in which the entire callus was used to initiate liquid cell cultures.

In order to determine whether sectors arose from a mixture of transformation events or are descendents of an epigenetic change following a single transformation

event, tissues from sectors of single calli were independently propagated as shown in the right column of figure 3-6A. The calli in the left column of figure 3-6A result from enlargement of the original microcalli picked from the transformation plates. Thus, the calli in the right column of figure 3-6A have not undergone the additional round of subculturing that was carried out for all other previously mentioned individual NT-1 lines (see methods). Calli from the sectors were used to initiate liquid suspension cultures and subjected to flow cytometry analysis (Figure 3-6B). In all cases, cultures derived from sectors, which by visual inspection had higher levels of GFP fluorescence, were demonstrated by flow cytometry to have a higher percentage of positive cells and/or higher levels of fluorescence in positive cells.

In order to determine if the sectors have the same genotype, DNA from the liquid suspension cultures arising from the sectors was isolated and subjected to Southern blot analysis (Figure 3-7). A 458 bp region common to all of the T-DNA constructs tested was used to probe genomic DNA cut with a methylation insensitive restriction enzyme, *Hin*dIII. Because this enzyme cuts only once in the T-DNA, the resulting fragments will be a unique size based on the next closest genomic *Hin*dIII site (unless a tandem T-DNA repeat flanks the right border). In four out of five sector pairs, the banding patterns match exactly (Figure 3-7, sector pairs 1, 2, 3, and 5) indicating a single transformation event. Sector pair 4, clearly has a different banding pattern suggesting this cell line may be composed of lineages representing two distinct transformation events. In any case, four out of five sector pairs demonstrate banding patterns consistent with identical genotypes. This is most likely an overestimate of the frequency of chimerism in populations used to

generate data in figures 3-3 and 3-4, since we investigated chimerism in only lines that displayed sectoring. Furthermore, the sectored calli were subjected to one fewer round of subculturing than the lines used to investigate the effects of MARs on transgene expression.

Discussion

While methods for using flow cytometry based on GFP expression in plants have been available for several years (Galbraith et al., 1995; Sheen et al., 1995) few applications have been reported: abscisic acid signal transduction in rice (Desikan et al., 1999; Hagenbeek et al., 2000; Gampala et al., 2001; Hagenbeek and Rock, 2001), root development in *Arabidopsis* (Birnbaum et al., 2003), flow cytometry protocol development and mosaic observations in tobacco leaves (Bohanec et al., 2002; Bastar et al., 2004), and nuclear targeting in tobacco (Grebenok et al., 1997). Only the last four cases contained examples of flow cytometry analysis of stably transformed plants and of these only the last three studies include attempts to quantify GFP expression. One of the true powers of flow cytometry is analysis of heterogeneity in a cell population, as properties are measured on an individual cell basis rather than as a population average (Yanpaisan et al., 1999). The data presented here represent an application of flow cytometry to characterize epigenetic variation that can exist in transgenic plant cell populations.

An important factor in our ability to quantitate GFP expression variegation in stably transformed cell lines is to accurately estimate the number of cells that have wild-

type levels of fluorescence. Due to the nature of the protoplasting procedure and the fact that in this study several samples had to be prepared simultaneously and rapidly, nuclei and protoplasts with compromised plasma membrane integrity frequently occurred in our preparations. Since GFP diffuses out of compromised protoplasts but they may otherwise have physical properties nearly indistinguishable from intact protoplasts on which data analysis is based, the number of cells with wild-type levels of fluorescence can be overestimated. In order to correct for this phenomenon, we used an approach common in animal systems by which cells are stained with propidium iodide to exclude necrotic cells from analysis, e.g. Migliaccio et al. (2000). Under conditions where debris is a problem, adding this step to existing plant cell flow cytometry protocols would enhance our ability to discriminate between expressing and non-expressing subpopulations of cells.

We have previously characterized the effects of flanking tobacco Rb7 MAR sequences on transgene expression in NT-1 tobacco suspension culture cells and whole tobacco plants (Allen et al., 1996; Ülker et al., 1999). In cultured cells, we observed a 60-fold increase in GUS expression by direct DNA transformation, but only a moderate 2 to 3-fold increase in whole plants. In our current study, we chose *Agrobacterium*-mediated transformation and observed a 2 to 4-fold increase in GFP expression in NT-1 tobacco suspension culture cells. There are several differences between our past and present experiments with MARs in NT-1 cells. The lower fold effects we have observed in this study may be due to any one or combination of these differences. *Agrobacterium*-mediated transformation results in less complex sites of integration than does direct DNA transformation. In the present work, the reporter gene and selectable marker reside on the

same vector, while they were on separate vectors in our previous work. We have used a different reporter gene and an assay that does not depend upon enzymatic activity of the transgene. Variation in mean expression values of control treatments may also account for fold differences. This is, when expression values for control constructs approach zero, the fold difference can dramatically increase as Cheng et al. (2001) noted when they determined the effect of the Rb7 MAR on GUS and GFP expression in rice.

The observed overall increase in GFP expression mediated by the Rb7 MAR was due in part to a greater proportion of cells that express GFP (i.e., cells in which GFP was not silenced). Additionally, the cells that express GFP express it at higher levels in MAR-containing lines than in control lines. Even in cell lines demonstrated not to be chimeric, we often observed epigenetic differences in the levels of GFP expression as evidenced by more than one mode of GFP expression level (Figure 3-6). Most often when more than one mode occurred the distribution was bimodal. On average across all treatments, 1.42 distinct epigenetic states per cell line were observed (differences between control and MAR-containing constructs were not significant ($p > 0.05$)).

In work we report here, we were able to assess epigenetic variation in transgene expression within isogenic cell lines. Because Brouwer et al. (2002) used traditional methods of assessing transgene expression in their studies of the effect of MARs on LUC expression in maize cells in culture, they were not able to assess epigenetic variation within a line. They did, however, examine variation across a large number of transgenic lines. The authors noted fewer non-expressing lines among lines that contained either the yeast ARS1 MAR or the maize *Adh1* 5' MAR than among control lines. Based on these

observations, they suggested that MARs increase the likelihood that an integrated gene will be expressed but not the level of expression (Brouwer et al., 2002). Similarly, Mankin et al. (2003) observed a MAR-induced reduction in the frequency of low expressing tobacco cell lines, but also noted that when these low expressing cell lines were removed from the analysis, MAR lines still showed an increase in GUS expression. Our data are in agreement with those of Brouwer et al. (2002) and Mankin et al. (2003) in that when MARs are incorporated into transgene constructs, we see an increase in the likelihood of expression in individual cells compared to whole plants or cell lines in the previous studies. While our data are consistent with those of Mankin et al. (2003) in that they observed greater expression in moderate to high expressing lines that contained MARs, it is unknown whether these increases in expression were due to greater proportions of expressing cells, cells expressing at higher levels, or both. In contrast to Brouwer et al. (2002), we observe that the expressing cells in lines transformed with the MAR-containing construct express at higher levels than the expressing cells in lines transformed with control constructs.

We set out to determine whether flow cytometry could be used to assess the effect of MARs on transgene expression in cultures derived from large pools of transformed cells and whether conclusions drawn from work with pools of transformed lines would parallel those drawn from work with individual cell lines. We found that flow cytometry of population cultures transformed with MAR-containing constructs resulted in unique profiles of GFP expression that were not present in control population cultures. These profiles are consistent with our observations on individual cell lines. In both cases,

presence of the Rb7 MAR increased the percentage of expressing cells as well as the levels of expression in expressing cells. Our results indicate that population cultures studied by flow cytometry may be a useful tool to quickly and easily evaluate the effect of various sequence elements on transgene expression.

While flow cytometry analysis allowed us to quantify differing modes of GFP expression within cell lines, it is possible that these differences could be attributed to chimeric cell populations. A few of the cell lines exhibited sectors that differed in GFP expression in calli cultured on solid media. Southern blot analysis demonstrated that in most cases the sectoring was due to epigenetic changes following an original transformation event. For example, in sector pair #5 (Figure 3-6) there was a complete loss of GFP expression, and this loss of expression was not accompanied by the loss of the T-DNA insert tested in southern blots (Figure 3-7). However, in one of the five sectored lines we tested, the original microcallus was most likely composed of cells from more than one transformation event. This observation stresses the need for additional rounds of subculturing to increase the probability of fixing a single genotype in each cell line. Our results in individual cell lines were based on one additional round of solid media subculture, which is expected to further minimize the frequency of chimeric cell lines.

In our present study we have not addressed the mechanism by which the Rb7 MAR increases GFP expression. However, we have observed that cells with the same genotype can include both GFP negative and GFP positive cells and that the Rb7 MAR reduces the frequency of GFP negative, non-expressing cells. This suggests MARs may

counteract the effects of gene silencing to increase the likelihood a transgene will be expressed.

Although it is generally accepted that MARs function only upon stable integration into the genome, minor increases in transient expression have been reported for the Rb7 MAR suggesting a possible classical enhancer effect (Allen et al., 1996; Cheng et al., 2001). Additionally, MARs may act to facilitate the activity of nearby enhancers as van der Geest et al. (1997) reported for the β-phaseolin 5' MAR. While it is not yet clear how the Rb7 MAR influences transgene expression, the fact that we observe higher levels of GFP expression in GFP positive cells compared to controls suggest a possible role in enhancer-like activity.

In summary, we have used flow cytometry to measure GFP expression in individual plant cells. Our results indicate that the Rb7 MAR increases the frequency of transgene expression, presumably by reducing gene silencing, while also increasing the levels of expression in expressing cells, perhaps through enhancer-like activity. These observations are consistent between individual isogenic cell lines and large pools of transformed cell lines.

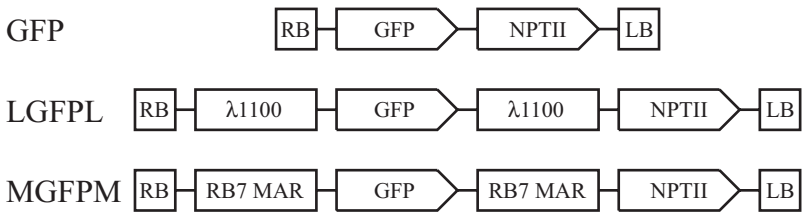
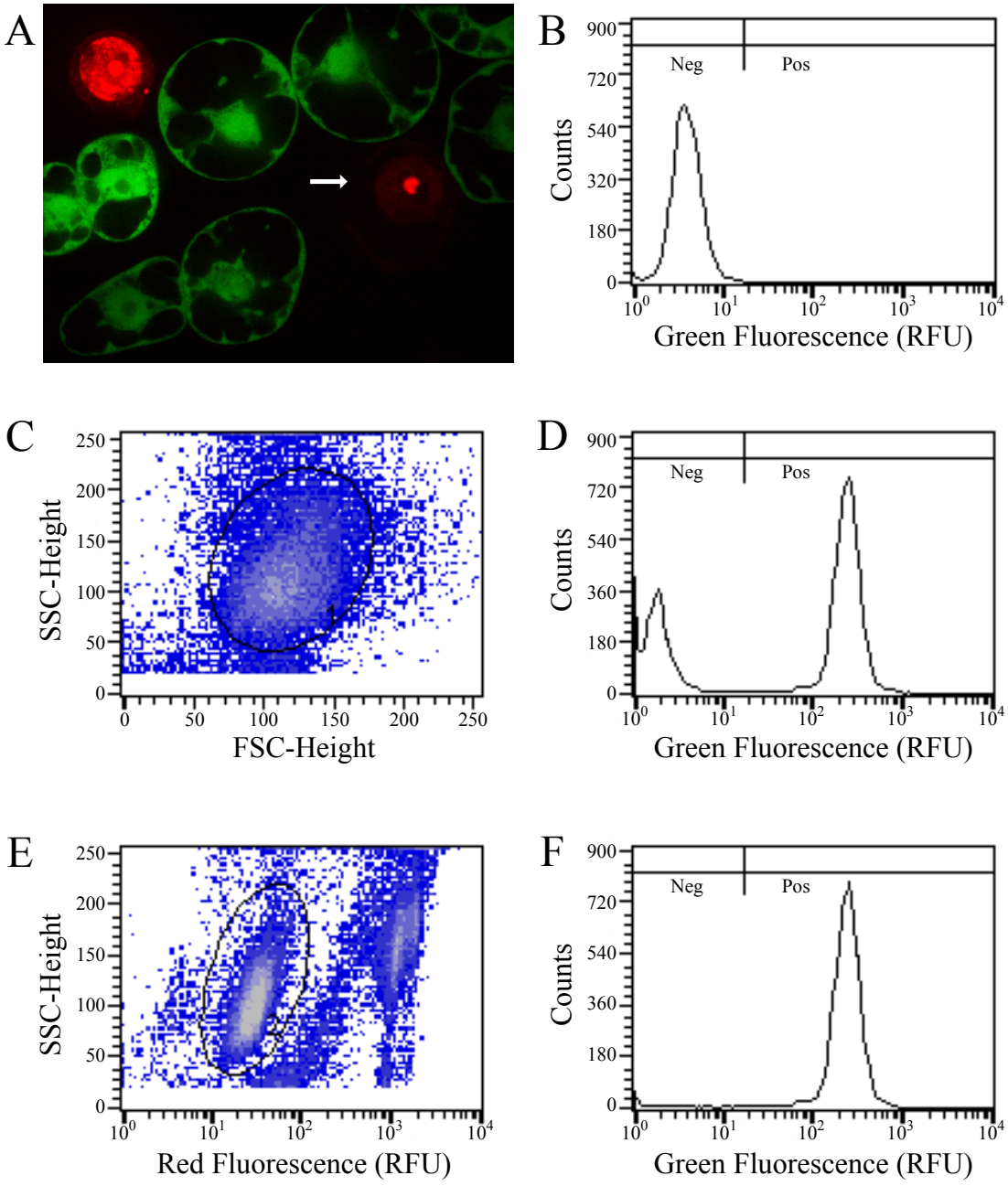


Figure 3-1. Schematic Diagrams of T-DNA Regions of Transformation Vectors.
The right and left boarders of the T-DNA are indicated with RB and LB, respectively. In the control (GFP) construct, the soluble-modified, red-shifted green fluorescent protein gene (GFP) is driven by the 35S promoter of CaMV with the nos polyadenylation region. The neomycin phosophotransferase selectable marker gene (NPTII) is driven by the nos promoter with the g7T polyadenylation region. An additional control construct (LGFPL) contains direct repeats of 1,195 bp of lambda phage spacer DNA. MGFPM contains direct repeats of the 1,167 bp Rb7 MAR. Diagram schematics are not to scale.

Figure 3-2. Propidium Iodide Staining of Protoplast Preparations Improves Flow Cytometry Resolution.
(A) A 1 m thick confocal section of a propidium iodide-stained protoplast preparation of a tobacco cell line uniformly expressing GFP after stable transformation with MGFPM. Intact protoplasts exclude propidium iodide (PI) while free nuclei and nuclei in protoplasts with compromised plasma membrane integrity stain brightly. GFP is absent from protoplasts in which nuclei stain brightly with PI. A white arrow indicates a missing portion of the plasma membrane. **(B)** Green fluorescence histogram of wild-type protoplasts. Fluorescence was measured in Relative Fluorescence Units (RFUs). Cells with green fluorescence below 17 RFUs are considered "GFP negative cells" (Neg) and the remaining cells are considered "GFP positive cells" (Pos). **(C)** MGFPM stably transformed protoplasts stained with PI subjected to biparametric analysis of side light scatter (SSC) versus forward light scatter (FSC). Gated events inside the oval are plotted in (D). **(D)** Green Fluorescence histogram of gated MGFPM protoplasts from (C). **(E)** The same data from the same MGFPM stably transformed protoplasts stained with PI were subjected to biparametric analysis of side light scatter (SSC) versus relative red fluorescence (PI). Gated events inside the oval are plotted in (F). These gated events have background levels of PI fluorescence similar to that of wild-type cells (data not shown). **(F)** Green fluorescence histogram of gated MGFPM protoplasts from (E). Note: after exclusion of protoplasts and debris that have high levels of PI fluorescence, the mode consistent with wild-type fluorescence was lost.


A
B
Neg
Pos
Counts
Green Fluorescence (RFU)
C
SSC-Height
FSC-Height
D
Neg
Pos
Counts
Green Fluorescence (RFU)
E
SSC-Height
Red Fluorescence (RFU)
F
Neg
Pos
Counts
Green Fluorescence (RFU)

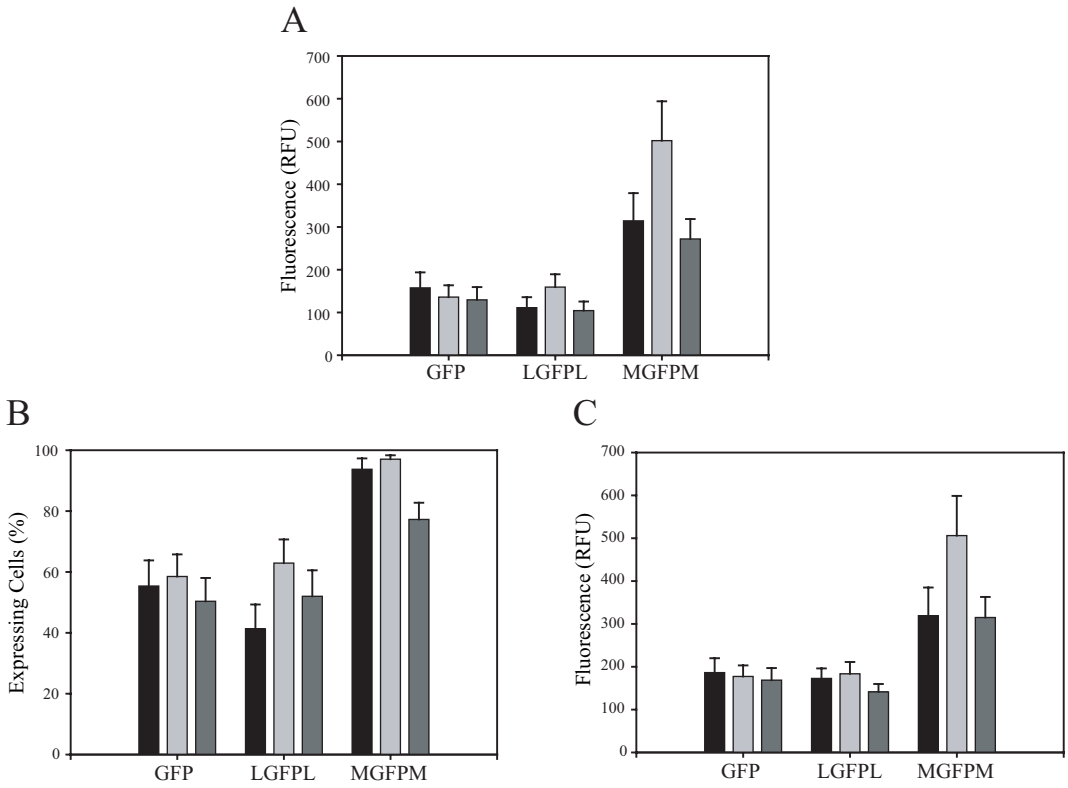


Figure 3-3. Average GFP Expression in Individual NT-1 Tobacco Cell Culture Lines for Three Replicate Experiments.

Tobacco protoplasts were analyzed by flow cytometry in three replicate experiments shown separately as black, light grey, and dark grey bars. Results from approximately 30 cell lines each of GFP, LGFPL, and MGFPM T-DNA vector transformations are shown for each replicate. **(A)** Total mean GFP fluorescence for cell lines transformed by each of the vectors (GFP positive cells and GFP negative cells are both included). **(B)** The mean percentage of GFP positive cells in lines transformed by each of the vectors. **(C)** Mean GFP fluorescence of GFP positive cells in cell lines transformed by each of the vectors. (Data from GFP negative cells is not included) Error bars denote standard errors of the mean.

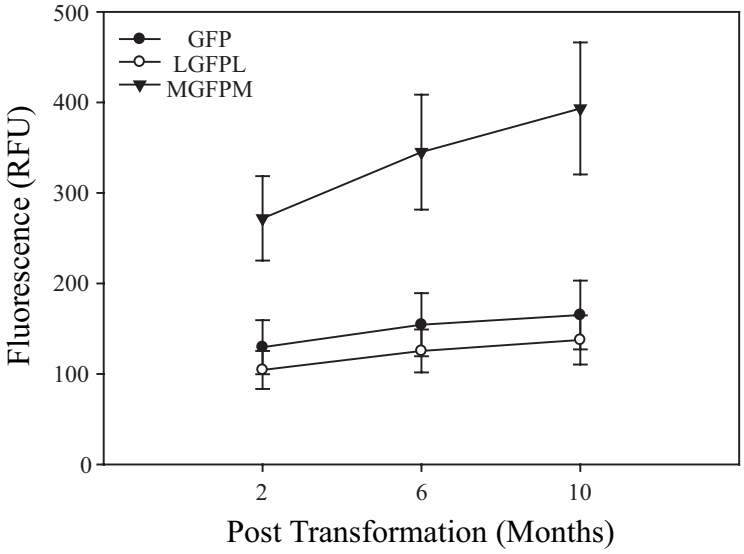


Figure 3-4. GFP Expression is Stable Over Time in Individual Cell Lines.
Cell lines were maintained on selection for ten months and green fluorescence was measured by flow cytometry. Error bars denote standard errors of the mean.

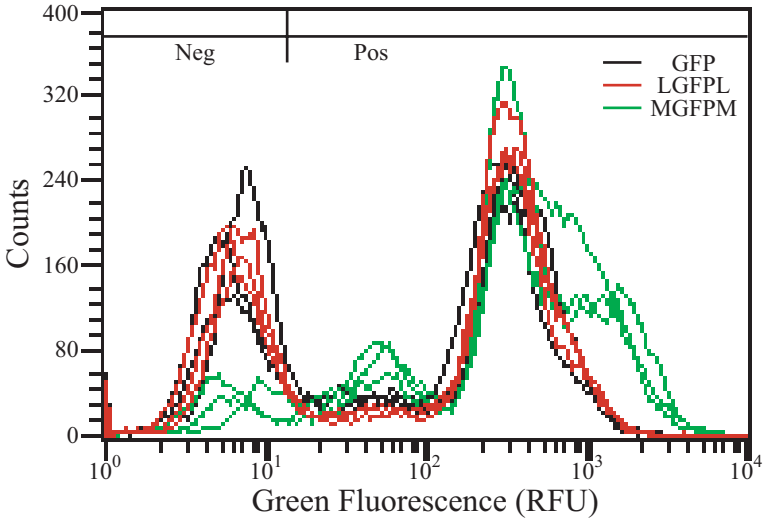
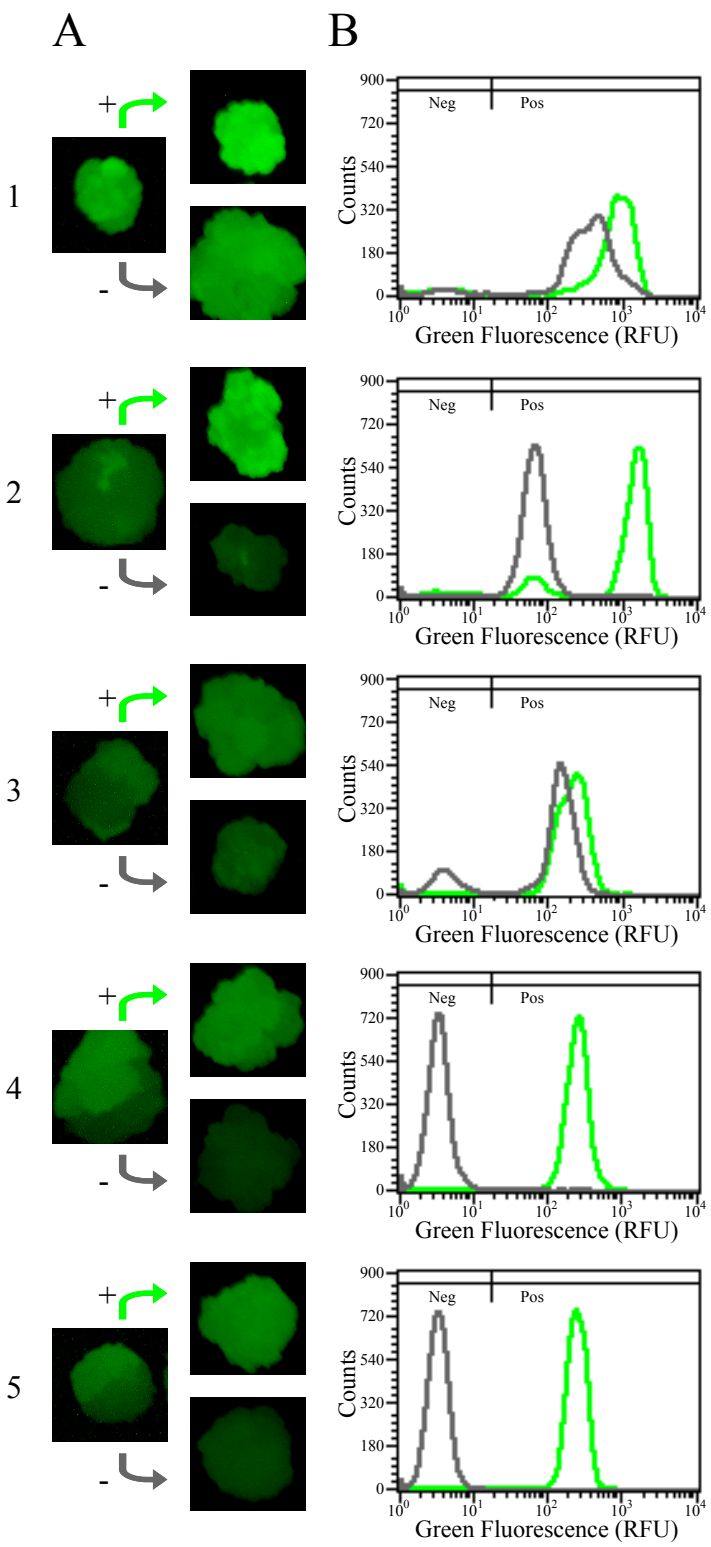


Figure 3-5. GFP Expression in Population NT-1 Tobacco Cell Cultures.
Green fluorescence histograms constructed from population cultures established from an entire Agrobacterium cocultivation transformation plate. Three independent cultures were established for each T-DNA vector. Parameters for histograms are the same as in figure 3-2B.

Figure 3-6. Isolation of Callus Sectors Expressing GFP at Different Levels and Flow Cytometric Analysis.
(A) A subset of calli display a sectored GFP phenotype (left column). Tissue from these sectors was subcultured on fresh solid medium (right column). (+) denotes tissue from the sector with the higher level of green fluorescence, while (-) denotes tissue from the sector with the lower level of green fluorescence. Sectored calli from each T-DNA vector were chosen for further analysis: sample #1 from GFP T-DNA; sample #2 from MGFPM T-DNA; samples #3, #4, and #5 from LGFPL T-DNA. **(B)** GFP fluorescence histograms of liquid cultures initiated from sectors in the right column of (A). The grey histogram is from the (-) sector and the green histogram is from the (+) sector. Parameters for histograms are the same as in figure 3-2B.


A
B
1
2
3
4
5
+
-
Neg
Pos
Counts
Green Fluorescence (RFU)
0
180
320
540
720
900

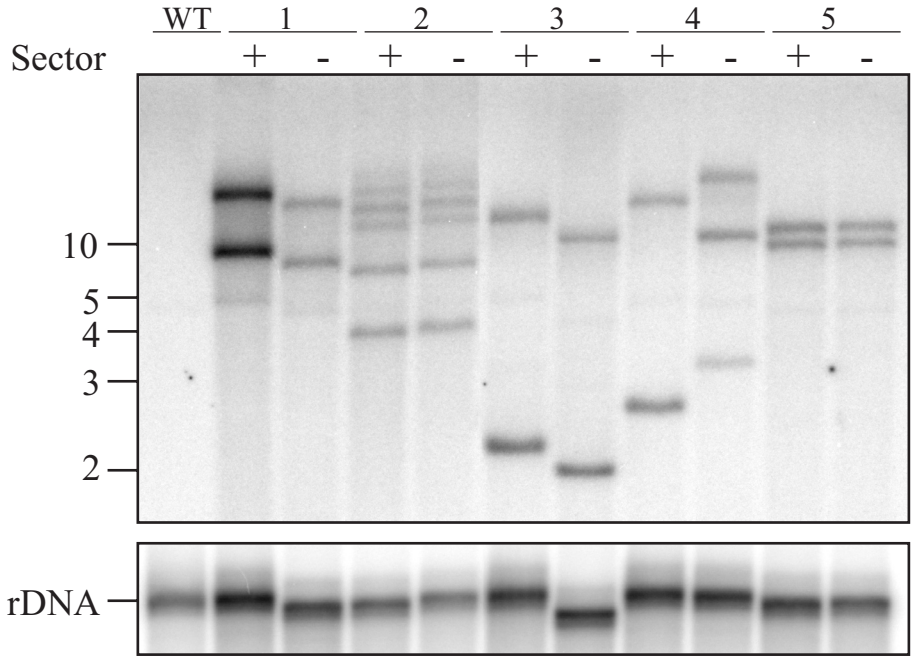


Figure 3-7. Genotyping Sectors by Southern Blot Analysis.
DNA was isolated from liquid cultures initiated from callus sectors (Figure 3-6A right column). Blots of 15µg of DNA digested with the methylation insensitive restriction enzyme *Hin*dIII (which cuts only once within the T-DNA) and separated by electrophoresis were probed with a portion of the right T-DNA border common to T-DNA in all vectors. Each banding pattern represents a transgenic fingerprint for each sector. The blot was reprobed with rDNA to account for differences in DNA loading and differences in DNA mobility attributed to possible differences in salt concentrations. Numbers to the left are approximate DNA size markers in kilobases.

References

Allen, G.C., Spiker, S., and Thompson, W.F. (2000). Use of matrix attachment regions (MARs) to minimize transgene silencing. Plant Mol Biol **43,** 361-376.

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

Allen, G.C., Hall, G., Jr., Michalowski, S., Newman, W., Spiker, S., Weissinger, A.K., and Thompson, W.F. (1996). High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. Plant Cell **8,** 899-913.

An, G.H. (1985). High-efficiency transformation of cultured tobacco cells. Plant Physiol **79,** 568-570.

Bastar, M.T., Luthar, Z., Skof, S., and Bohanec, B. (2004). Quantitative determination of mosaic GFP gene expression in tobacco. Plant Cell Rep **22,** 939-944.

Becker, D., Kemper, E., Schell, J., and Masterson, R. (1992). New plant binary vectors with selectable markers located proximal to the left T-DNA border. Plant Mol Biol **20,** 1195-1197.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science **302,** 1956-1960.

Bode, J., Stengert-Iber, M., Kay, V., Schlake, T., and Dietz-Pfeilstetter, A. (1996). Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. Crit Rev Eukaryot Gene Expr **6,** 115-138.

Bohanec, B., Luthar, Z., and Rudolf, K. (2002). A protocol for quantitative analysis of green fluorescent protein-transformed plants, using multiparameter flow cytometry with cluster analysis. Acta Biologica Cracoviensia Series Botanica **44,** 145-153.

Bonifer, C., Vidal, M., Grosveld, F., and Sippel, A.E. (1990). Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. Embo J **9,** 2843-2848.

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

Brouwer, C., Bruce, W., Maddock, S., Avramova, Z., and Bowen, B. (2002). Suppression of transgene silencing by matrix attachment regions in maize: a dual role for the maize 5' ADH1 matrix attachment region. Plant Cell **14,** 2251-2264.

Callaway, A., Liu, W., Andrianov, V., Stenzler, L., Zhao, J., Wettlaufer, S., Jayakumar, P., and Howell, S.H. (1996). Characterization of cauliflower mosaic virus (CaMV) resistance in virus-resistant ecotypes of Arabidopsis. Mol Plant Microbe Interact **9,** 810-818.

Carrington, J.C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. Science **301,** 336-338.

Chakravarti, I.M., Laha, R.G., and Roy, J. (1967). Handbook of methods of applied statistics. (New York,: Wiley).

Cheng, Z.Q., Targolli, J., and Wu, R. (2001). Tobacco matrix attachment region sequence increased transgene expression levels in rice plants. Mol Breeding **7,** 317-327.

Davis, S.J., and Vierstra, R.D. (1998). Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol Biol **36,** 521-528.

De Bolle, M.F.C., Butaye, K.M.J., Coucke, W.J.W., Goderis, I.J.W.M., Wouters, P.F.J., van Boxel, N., Broekaert, W.F., and Cammue, B.P.A. (2003). Analysis of the influence of promoter elements and a matrix attachment region on the inter-individual variation of transgene expression in populations of Arabidopsis thaliana. Plant Sci **165,** 169-179.

Desikan, R., Hagenbeek, D., Neill, S.J., and Rock, C.D. (1999). Flow cytometry and surface plasmon resonance analyses demonstrate that the monoclonal antibody JIM19 interacts with a rice cell surface component involved in abscisic acid signalling in protoplasts. FEBS Lett **456,** 257-262.

Galbraith, D.W., Herzenberg, L.A., and Anderson, M.T. (1999a). Flow cytometric analysis of transgene expression in higher plants: green fluorescent protein. Methods Enzymol **302,** 296-315.

Galbraith, D.W., Anderson, M.T., and Herzenberg, L.A. (1999b). Flow cytometric analysis and FACS sorting of cells based on GFP accumulation. Methods Cell Biol **58,** 315-341.

Galbraith, D.W., Lambert, G.M., Grebenok, R.J., and Sheen, J. (1995). Flow cytometric analysis of transgene expression in higher plants: green-fluorescent protein. Methods Cell Biol **50,** 3-14.

Gampala, S.S., Hagenbeek, D., and Rock, C.D. (2001). Functional interactions of lanthanum and phospholipase D with the abscisic acid signaling effectors VP1 and ABI1-1 in rice protoplasts. J Biol Chem **276,** 9855-9860.

Grebenok, R.J., Lambert, G.M., and Galbraith, D.W. (1997). Characterization of the targeted nuclear accumulation of GFP within the cells of transgenic plants. Plant J **12,** 685-696.

Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science **301,** 798-802.

Hagenbeek, D., and Rock, C.D. (2001). Quantitative analysis by flow cytometry of abscisic acid-inducible gene expression in transiently transformed rice protoplasts. Cytometry **45,** 170-179.

Hagenbeek, D., Quatrano, R.S., and Rock, C.D. (2000). Trivalent ions activate abscisic acid-inducible promoters through an ABI1-dependent pathway in rice protoplasts. Plant Physiol **123,** 1553-1560.

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

Han, K.H., Ma, C.P., and Strauss, S.H. (1997). Matrix attachment regions (MARs) enhance transformation frequency and transgene expression in poplar. Transgenic Res **6,** 415-420.

Holmes-Davis, R., and Comai, L. (2002). The matrix attachment regions (MARs) associated with the Heat Shock Cognate 80 gene (HSC80) of tomato represent specific regulatory elements. Mol Genet Genomics **266,** 891-898.

Johnson, E., Miklas, P.N., Stavely, J.R., and Martinezcruzado, J.C. (1995). Coupling-phase and repulsion-phase Rapds for marker assisted selection of Pi-181996 rust resistance in common bean. Theor Appl Genet **90,** 659-664.

Liu, J.W., and Tabe, L.M. (1998). The influences of two plant nuclear matrix attachment regions (MARs) on gene expression in transgenic plants. Plant Cell Physiol **39,** 115-123.

Mankin, S.L. (2000). Chromatin elements and transgene expression in tobacco cells [dissertation]. In Botany (Raleigh, NC: North Carolina State University).

Mankin, S.L., Allen, G.C., Phelan, T., Spiker, S., and Thompson, W.F. (2003). Elevation of transgene expression level by flanking matrix attachment regions (MAR) is promoter dependent: a study of the interactions of six promoters with the RB7 3' MAR. Transgenic Res **12,** 3-12.

Martelli, A.M., Falcieri, E., Zweyer, M., Bortul, R., Tabellini, G., Cappellini, A., Cocco, L., and Manzoli, L. (2002). The controversial nuclear matrix: a balanced point of view. Histol Histopathol **17,** 1193-1205.

Matzke, M., and Matzke, A.J. (2003). RNAi extends its reach. Science **301,** 1060-1061.

McGann, L.E., Walterson, M.L., and Hogg, L.M. (1988). Light scattering and cell volumes in osmotically stressed and frozen-thawed cells. Cytometry **9,** 33-38.

Mendu, N., Massel, M., and Spiker, S. (2001). Increasing loop domain size does not diminish effects of matrix attachment regions on transgene expression in tobacco cells culture. FEBS Lett **496,** 66-67.

Migliaccio, A.R., Bengra, C., Ling, J., Pi, W., Li, C., Zeng, S., Keskintepe, M., Whitney, B., Sanchez, M., Migliaccio, G., and Tuan, D. (2000). Stable and unstable transgene integration sites in the human genome: extinction of the green fluorescent protein transgene in K562 cells. Gene **256,** 197-214.

Mlynarova, L., Jansen, R.C., Conner, A.J., Stiekema, W.J., and Nap, J.P. (1995). The MAR-mediated reduction in position effect can be uncoupled from copy number-dependent expression in transgenic plants. Plant Cell **7,** 599-609.

Mlynarova, 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.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res **8,** 4321-4325.

Phi-Van, L., von Kries, J.P., Ostertag, W., and Stratling, W.H. (1990). The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. Mol Cell Biol **10,** 2302-2307.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning : a laboratory manual. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

Schramke, V., and Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science **301,** 1069-1074.

Shapiro, H.M. (1995). Practical flow cytometry. (New York: Wiley-Liss).

Sheen, J., Hwang, S., Niwa, Y., Kobayashi, H., and Galbraith, D.W. (1995). Green-fluorescent protein as a new vital marker in plant cells. Plant J **8,** 777-784.

Sidorenko, L., Bruce, W., Maddock, S., Tagliani, L., Li, X., Daniels, M., and Peterson, T. (2003). Functional analysis of two matrix attachment region (MAR) elements in transgenic maize plants. Transgenic Res **12,** 137-154.

Snedecor, G.W., and Cochran, W.G. (1989). Statistical methods. (Ames: Iowa State University Press).

Stief, A., Winter, D.M., Stratling, W.H., and Sippel, A.E. (1989). A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature **341,** 343-345.

Ülker, B., Allen, G.C., Thompson, W.F., Spiker, S., and Weissinger, A.K. (1999). A tobacco matrix attachment region reduces the loss of transgene expression in the progeny of transgenic tobacco plants. Plant J **18,** 253-263.

Vain, P., Worland, B., Kohli, A., Snape, J.W., Christou, P., Allen, G.C., and Thompson, W.F. (1999). Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J **18,** 233-242.

van der Geest, A.H., and Hall, T.C. (1997). The beta-phaseolin 5' matrix attachment region acts as an enhancer facilitator. Plant Mol Biol **33,** 553-557.

van Leeuwen, W., Mlynarova, L., Nap, J.P., van der Plas, L.H., and van der Krol, A.R. (2001). The effect of MAR elements on variation in spatial and temporal regulation of transgene expression. Plant Mol Biol **47,** 543-554.

Weiler, K.S., and Wakimoto, B.T. (1995). Heterochromatin and gene expression in Drosophila. Annu Rev Genet **29,** 577-605.

Yanpaisan, W., King, N.J.C., and Doran, P.M. (1999). Flow cytometry of plant cells with applications in large-scale bioprocessing. Biotechnol Adv **17,** 3-27.

APPENDIX A

Flow Cytometric Analysis of GFP Expression Variegation Mediated by a Repetitive DNA Sequence (RPS) and Interactions with the Rb7 MAR

Plants have endogenous repetitive sequences that may be organized into transcriptionally inactive heterochromatin. If by chance a transgene integrates near one of these plant sequences, silencing events may hinder the expression of the transgene (Prols and Meyer, 1992). Previously, ten Lohuis et al. (1995) studied the effect of a repetitive DNA sequence (RPS) from *Petunia hybrida* (10^4 copies/ haploid genome) on expression of a nearby transgene. In both *Petunia hybrida* and *Nicotiana tabacum* including the RPS element in transgene constructs resulted in variegated GUS expression and served as a functioned as a hot spot for *de novo* DNA methylation (a hallmark of transcriptionally inactive chromatin). Not only is the RPS element repetitive throughout the genome, but it also contains several inverted repeats within each element that may be signals for hypermethylation (Muller et al., 2002).

Given the observation that the RPS element induces transgene expression variegation, we chose to use flow cytometry measurement of GFP expression to quantify this variegation. Also, since the Rb7 MAR was shown in chapter three of this dissertation to alter variegation, we chose to explore whether the Rb7 MAR could counteract the effects of the RPS element. Using *Agrobacterium*-mediated transformation, tobacco NT-1 cells in culture were transformed with the T-DNA constructs in figure A-1. All methods for transformation, cell culture and flow cytometry

analysis were conducted in the exact manner described in the methods section of chapter three. In fact experiments with RPS constructs were carried out side-by-side with replicate three described in chapter 3. For simplicity the data from replicate three is reproduced in this appendix.

Flow cytometric analysis from all cells, both GFP negative and GFP positive, are included in figure A-2A. This type of analysis ignores transgene expression variegation. Unexpectedly, expression levels in RPS containing constructs (RGFP, RLGFP, and RMGFP) were unusually high. In fact, cell lines transformed with either RGFP or RMGFP had significantly higher GFP fluorescence ($p < 0.05$) when compared to lines carrying GFP alone. Cell lines transformed with RLGFP had borderline significant ($p = 0.075$) higher GFP fluorescence. The RPS containing lines were not significantly different from each other ($p > 0.05$). These surprising results suggest that the RPS element in our hands actually increases transgene expression. Similar to our analysis of the Rb7 MAR, the question remained: Were these increases due to a higher percentage of cells expressing the transgene, higher expression in expressing cells, or both?

In order to address the first possibility, the percentage of GFP positive cells was determined for each cell line. The averages of these percentages for all cell lines of each treatment are shown in figure A-2B. Cell lines transformed with either RGFP, RLGFP, or RMGFP had significantly higher percentages of positive cells ($p < 0.05$) when compared to lines carrying GFP alone. All of the RPS containing lines were not significantly different from each other ($p > 0.05$). These results suggest that the RPS element actually

decreases variegation associated with non-expressing cells by increasing the likelihood of GFP expression.

In order to address the second possibility that expression levels of individual expressing cells are higher in lines containing the RPS element, data were plotted for positive cells only. The averages of the magnitudes of GFP expression of expressing cells for cell lines of each treatment are shown in figure A-2C. Even though the average magnitudes of GFP expression in the lines transformed with the RPS element were higher than those transformed with GFP alone, none of the differences were significant ($p > 0.05$). Again, all of the RPS containing lines were not significantly different from each other ($p > 0.05$). These results do not support the suggestion that the RPS element increases the magnitude of GFP expression in expressing cells.

In order to evaluate possible changes in GFP expression over time, cell lines were maintained with kanamycin selection after the first flow cytometry analysis was conducted. After four additional months (with weekly passages), green fluorescence data was collected. No statistically significant changes in mean GFP expression, percentages of expressing cells, or the levels of GFP expression in expressing cells were observed for any of the treatments (data not shown). Additionally, a subset of cell lines in each treatment were cultured without kanamycin selection for four months. Again, no changes in mean GFP expression levels or percentages of expressing cells were observed for any of the treatments (data not shown).

Contrary to previous observations of increased transgene expression variegation when the RPS element was present near transgenes in both *Petunia hybrida* and

Nicotiana tabacum (ten Lohuis et al., 1995), we actually observed reduced transgene expression variegation associated with non-expressing cells. In addition, the RPS element actually increased overall GFP expression. This increase was due to a greater percentage of expressing cells, but not due to an increase in the magnitude of expression in expressing cells. These results suggest the RPS element increases the likelihood of GFP expression.

A potential problem may have arisen from using a selectable marker that is physically linked to the RPS element. Regenerating transformants on selection may bias for a locus that is poised for expression and not subject to effects by the RPS element, although in previous experiments, ten Lohuis et al. (1995) flanked the RPS element by the selectable marker and the reporter gene and still observed variegated GUS expression. In an attempt to limit the effect of the RPS element on the selectable marker, we chose to place the selectable marker at the opposite T-DNA border for the greatest degree of separation from the RPS element. It remains unknown whether this difference would explain the results we observed.

Currently it is unknown whether the RPS element has directionality, but in our T-DNA vectors, the RPS element was inserted in the inverse orienatation compared to ten Lohuis et al. (1995). This may have also contributed to our unexpected results. Recently, others have observed unexpected results. Muller et al. (2002) did not observe variegated expression in *Arabidopsis thaliana* when the RPS element was included in transgene constructs. The authors did, however, observe hypermethylation within repeat sequences of the RPS element.

Surprisingly, our results suggest the RPS element can increase levels of transgene expression in a population by increasing the likelihood of transgene expression. A possible explanation for our observations may be that the RPS element may have caused variegation in selectable marker expression. In this case, selection bias may eliminate cells that had the capacity to be negatively influenced by the RPS element, thus reducing the number of cells with low expression. We removed a subset of cell lines in liquid culture from selection, but did not notice any differences. It's possible that variegation may be established early in NT-1 calli on solid media as was observed in chapter 3, and removal of kanamycin selection at these earlier times may have yielded different results. It should also be noted, that no gross differences in transformation efficiencies were observed in treatments containing the RPS element (data not shown).

Further investigation will be required to understand how the RPS element influences the expression of neighboring genes. One possible approach would be to avoid selection bias by PCR screening for transformed *Arabidopsis* plants. This approach has proven successful in identifying T-DNA insertions into regions of the genome that are under-represented in T-DNA libraries where antibiotic selection is required for identification (Francis, 2004).

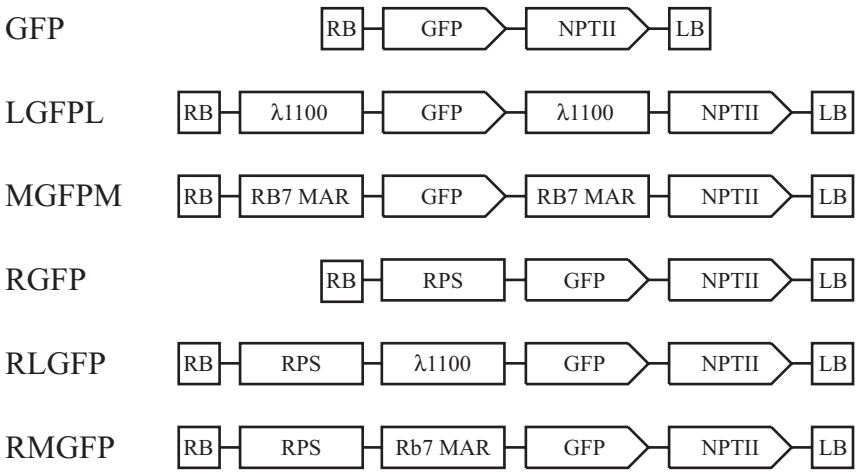


Figure A-1. Schematic Diagrams of T-DNA Regions of Transformation Vectors.
The right and left boarders of the T-DNA are indicated as RB and LB, respectively. In the control (GFP) construct, the soluble-modified, red-shifted green fluorescent protein gene (GFP) is driven by the 35S promoter of CaMV with the nos polyadenylation region. The neomycin phosophotransferase selectable marker gene (NPTII) is driven by the nos promoter with the g7T polyadenylation region. Construct (LGFPL) contains direct repeats of 1,195 bp of lambda phage spacer DNA. MGFPM contains direct repeats of the 1,167 bp Rb7 MAR. The 1,553 bp repetitive DNA sequence (RPS) was inserted upstream of the 35S promoter in three test constructs (RGFP, RLGFP, and RMGFP) with or without an intervening DNA sequence. The same lambda phage spacer DNA is used in RLGFP as in LGFPL. RMGFP contains the same Rb7 MAR sequence as MGFPM. Diagram schematics are not to scale. The cloning of GFP, LGFPL, and MGFPM was described in chapter 3. RGFP was kindly provided by Luke Mankin and originally described as pLMNC-b49 (Mankin, 2000). RLGFP and RMGFP were kindly provided by Anton Callaway and originally described as pAC07 and pAC10 respectively.

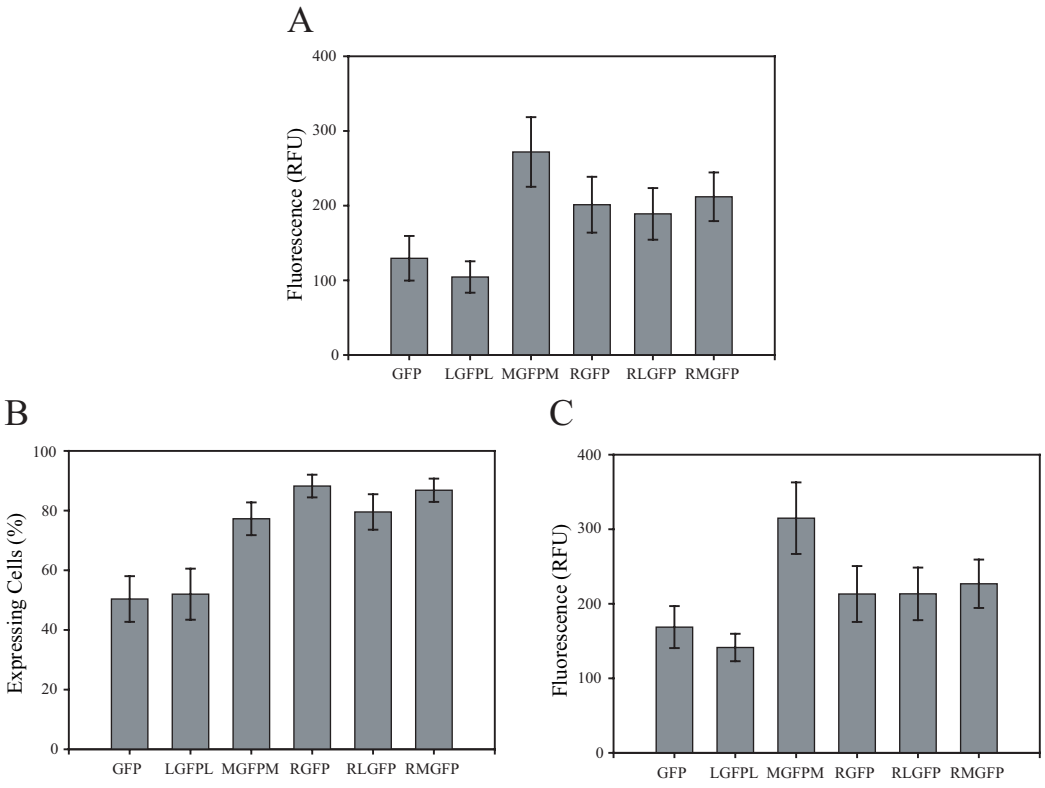


Figure A-2. Average GFP Expression in Individual NT-1 Tobacco Cell Culture Lines. Tobacco protoplasts were analyzed by flow cytometry. Results from approximately 30 cell lines each of GFP, LGFPL, MGFPM, RGFP, RLGFP, and RMGFP T-DNA vector transformations are shown. Data from GFP, LGFPL, and MGFPM were previously reported in chapter 3 and re-displayed here for convenience. **(A)** Total mean GFP fluorescence for cell lines transformed by each of the vectors (GFP positive cells and GFP negative cells are both included). **(B)** The mean percentage of GFP positive cells in lines transformed by each of the vectors. **(C)** Mean GFP fluorescence of GFP positive cells in cell lines transformed by each of the vectors. (Data from GFP negative cells is not included) Error bars denote standard errors of the mean.

References

Francis, K.E. (2004). The effects of T-DNA integration sites on transgene expression in Arabidopsis [dissertation]. In Genetics (Raleigh, NC: North Carolina State University).

Mankin, S.L. (2000). Chromatin elements and transgene expression in tobacco cells [dissertation]. In Botany (Raleigh, NC: North Carolina State University).

Muller, A., Marins, M., Kamisugi, Y., and Meyer, P. (2002). Analysis of hypermethylation in the RPS element suggests a signal function for short inverted repeats in de novo methylation. Plant Mol Biol **48,** 383-399.

Prols, F., and Meyer, P. (1992). The methylation patterns of chromosomal integration regions influence gene activity of transferred DNA in Petunia hybrida. Plant J **2,** 465-475.

ten Lohuis, M., Muller, A., Heidmann, I., Niedenhof, I., and Meyer, P. (1995). A repetitive DNA fragment carrying a hot spot for de novo DNA methylation enhances expression variegation in tobacco and petunia. Plant J **8,** 919-932.