Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 3490–3496, April 1994

Review

Inactivation of gene expression in plants as a consequence of specific sequence duplication

R. B. Flavell

John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom

ABSTRACT **Numerous examples now exist in plants where the insertion of multiple copies of a transgene leads to loss of expression of some or all copies of the transgene. Where the transgene contains sequences homologous to an endogenous gene, expression of both transgene and endogenous gene is sometimes found to be impaired. Several examples of these phenomena displaying different features are reviewed. Possible explanations for the observed phenomena are outlined, drawing on known cellular processes in *Drosophila*, fungi, and mammals as well as plants. It is hypothesized that duplicated sequences can, under certain circumstances, become involved in cycles of hybrid chromatin formation or other processes that generate the potential for modification of inherited chromatin structure and cytosine methylation patterns. These epigenetic changes could lead to altered transcription rates or altered efficiencies of mRNA maturation and export from the nucleus. Where the loss of gene expression is posttranscriptional, antisense RNA could be formed on accumulated, inefficiently processed RNAs by an RNA-dependent RNA polymerase or from a chromosomal promoter and cause the observed loss of homologous mRNAs and possibly the modification of homologous genes. It is suggested that the mechanisms evolved to help silence the many copies of transposable elements in plants. Multicopy genes that are part of the normal gene catalog of a plant species must have evolved to avoid these silencing mechanisms or their consequences.**

Frequently, when unexpected phenomena are observed they are ignored, then timidly explored and discussed, and only later published and debated with firm conviction in a more coherent framework. We have now entered the third phase for the unexpected phenomena associated with the silencing of multiple copies of genes inserted into plants and the silencing of endogenous plant genes with sequence homology to the newly inserted DNA. The discoveries are extremely important because (*i*) they reveal previously unrecognized facets of the control of gene expression except as rare isolated instances; (*ii*) they raise serious questions for all those wishing to exploit transgenic plants in research laboratories, industries, or agriculture; and (*iii*) they have opened up avenues of research in plant biology, including the means of making mutants.

The series of reports that established the phenomenon were published from 1989 to 1991, some 7 years after the first series of publications announcing the production of transgenic plants (but see refs. 1 and 2). Prominent early publications were those that showed that insertion of an additional copy of a chalcone synthase or dihydroflavonol-4-reductase gene into petunia plants led to the silencing in many but not all the transgenic plants of the inserted gene and its endogenous homologues (3–5). The coordinate silencing of the transgene and the homologous endogenous gene gave rise to the term "cosuppression" (3, 5).

Chalcone synthase facilitates the conversion of coumaryol CoA and 3-malonyl CoA to chalcone in the pathway of anthocyanin pigment biosynthesis. One gene of the small multigene family of chalcone synthase in petunia is especially active in petunia flowers, where pigment production in the corolla and anthers is substantial. After insertion of the chalcone synthase gene under the control of the "constitutive" cauliflower mosaic virus (CaMV) 35S promoter, up to 50% of the transgenic plants with the new gene showed sectors of reduced or no anthocyanin pigment in the flowers, and in some plants the flowers were completely white (3). This lack of pigment is correlated with very low levels of mRNA from the newly inserted and the endogenous copies of the chalcone synthase genes. In a similar but not identical series of plants Mol and coworkers (6, 7) found from nuclear "run on" experiments that nascent RNA transcription is unaffected. I have confirmed this (unpublished data).

In some plants whose flowers are white due to the inserted chalcone synthase gene, occasional branches occur with purple flowers or flowers with purple sectors. All the flowers on such a branch are usually very similar in pattern compared with the flowers on other branches of the plant. Thus, it can be concluded that a somatically inherited change of state of the transgene has occurred in this branch and, in some cases, during formation of the meristem of the branch. The occurrence of such epigenetic changes during plant or flower development emphasizes that the silencing phenomenon is reversible and can be under developmental control (5, 8). The new phenotype of a flower (and of the branch containing the flower) can be inherited (9) through sexual reproduction and seed development/germination. Many different "states" of transgene activity in cosuppression are possible based on the variety of flower pigmentation patterns arising in isogenic progeny. The patterns of floral pigmentation imply that the capacity of the transgene to cause suppression is subject to developmental influences.

This example illustrates many of the features of cosuppression but many other examples are now known in six plant species where gene inactivation results from the introduction of homologous sequences (reviewed in refs. 10 and 11). Some of these are noted below.

Tomato plants transformed with a truncated gene encoding part of polygalacturonase under the control of the CaMV 35S promoter showed only low levels of endogenous polygalacturonase mRNA in ripening fruit where it is normally highly induced. Transcripts of the transgene were similarly reduced but only in fruit (12). Thus, transcription of both genes may be essential for mutual inhibition. Degraded RNA products from both genes were found in cosuppressed fruit, suggesting that RNA transcription is not inhibited and therefore loss of mature mRNA is due to posttranscriptional turnover.

Multiple copies of transgenes can interact to cosuppress each other (13, 14). Some *Arabidopsis thaliana* plants carrying multiple closely linked copies of the hygromycin phosphotransferase gene lost resistance to hygromycin during development (14). This multicopy locus also suppressed active copies of the same gene at another locus introduced by crossing two strains. The cosuppression

Abbreviations: CaMV, cauliflower mosaic virus; SAR, scaffold attachment region.

could be reversed by separating the loci by outcrossing. These results suggested that the initial silencing of the multiple inserts at one locus, as well as the silencing/activation of the genes in different loci, is the consequence of specific sequence homology-dependent interactions.

Similar results were found when a portion of the nopaline synthase gene under the control of the CaMV 35S promoter was introduced into tobacco plants already containing wild-type nopaline synthase genes (15). All progeny containing both introduced genes expressed no nopaline synthase activity. The lack of nopaline synthase activity was dependent on the presence of the portion of the nopaline synthase gene in the genes introduced second. When the wild-type nopaline synthase gene was segregated away from the duplicated nopaline synthase sequence its activity was fully or partially restored in most plants. However, suppression of activity was not correlated with methylation of cytosines in the promoter as found in other cases (2, 16).

Vaucheret (17) has described an example of a transgene inserted into tobacco that contains a chimeric gene of nitrite reductase in the antisense direction under the control of the CaMV 35S promoter linked to a sequence encoding neomycin phosphotransferase under the control of the CaMV 19S promoter. No expression of either gene could be detected and this transgene complex suppresses any gene under the control of the 19S and 35S promoter inserted by sexual crossing. Thus, the transgene complex is a strong cosuppressor of genes containing related sequences. Cosuppression was stable but not observed until transient expression of newly introduced transferred DNA (T-DNA) had subsided—i.e., cosuppression probably is a postintegration event (but see ref. 16). Ninety base pairs of homology in promoter sequences were sufficient to create a cosuppressed condition.

These kinds of results help explain and were reinforced by other examples in which transgenes present in one copy in a plant were much more active than transgenes present in two or more copies (1, 16, 18–21).

In another well-investigated example, transgenic tobacco plants were created that contained two introduced genes, each conferring resistance to a different antibiotic (2, 22–25). The coding sequences of the two genes were unrelated in sequence but they had a common promoter contained in two copies of 300 bp taken from the nopaline synthase gene of *Agrobacterium tumefaciens*. One gene was introduced at the first transformation step and the other was introduced at a second transformation step. Double transformants were produced readily but, surprisingly, in 15% of the double transformants the phenotype of the gene conferring kanamycin resistance (introduced first) was lost after introduction of the second gene conferring hygromycin resistance. In this case though, in contrast to the previously mentioned examples, there was no cosuppression in that the plants that had lost kanamycin resistance retained hygromycin resistance. (This could have been the result of selection for hygromycin resistance in the second transformation step.) Loss of kanamycin resistance was accompanied by methylation of specific CpG residues in its nopaline synthase promoter sequence. The dependence of the trans-inactivation on the second gene encoding hygromycin resistance was illustrated by at least partial reactivation of the kanamycin-resistance gene and loss of cytosine methylation in its promoter when the hygromycin gene was segregated away from the kanamycin-resistance gene.

When the antibiotic-resistance transgenes were combined by sexual crossing, instead of by transformation, similar inactivation of the kanamycin-resistance phenotype in the presence of the hygromycin-resistance gene was recorded, thus eliminating the possibility that physiological states peculiar to the transformation procedures were the cause of kanamycin gene inactivation. Independently integrated kanamycin-resistant transgenes whose activity could not be suppressed by specific hygromycin-resistance transgenes introduced by a second transformation event also could not be inactivated by the bringing together of the two introduced genes by sexual crossing. Different transgene combinations produced no, partial (unstable), or complete (stable) trans-inactivation of kanamycin resistance. These results imply that the ability to trans-inactivate or to be trans-inactivated is defined by the state of the gene loci, including possibly their position in the chromosome and/or in the nucleus. Of special interest is the observation that in some plants homozygous for kanamycin-resistance genes, complete somatically stable trans-inactivation of kanamycin resistance occurred more readily than in plants heterozygous for the kanamycin-resistance genes (see also ref. 26).

Assessment of all these examples indicates that cosuppression and trans-inactivation of genes are dependent on sequence homology, can be epigenetically reversible, break down to various extents when the homologous loci are segregated away from one another, are sometimes dependent on transcription or a promoter, are sometimes under developmental control, do not occur with all copies of transgenes, and are sometimes associated with changes in the cytosine methylation pattern of genes. In some examples, the inactivation is reciprocal—i.e., all copies are inactivated while in others only one copy is inactivated. Homozygous transgenes are sometimes silenced more effectively than hemizygous transgenes. Loss of gene expression appears to be due to inhibition of transcription in some cases and degradation of mRNA in others. This plethora of unexpected phenomena needs to be explained.

The examples quoted above have emerged from recent research involving transgenes. However, a few examples of allelic interactions leading to silenced states of endogenous genes, some of which are inherited, have been known in plants for a long time (9–11). These examples include paramutation, in which one special allele (the paramutable allele) is converted to a state of activity displayed by the other allele (the paramutagenic allele). The new state is inherited but is often epigenetically unstable and may revert after the paramutable allele has been separated from the paramutagenic allele in progeny segregation (27–29).

Studies of the *R* locus in maize, which confers pigmentation in the aleurone layer of the seed, have shown that paramutable alleles are inherently unstable even in the absence of a paramutagenic allele. This suggests that gene structure is an important element of the potential to interact with a homologous allele. The developmental stage when the genetic change occurs and the subsequent stability of the paramutated allele also differ from one paramutable allele to another (29). Meyer *et al.* (30) discovered paramutagenic-like versions of the maize *A1* gene inserted into petunia under the control of the CaMV 35S promoter. Epigenetic variants of an unstable transgene were characterized. One homozygous variant had lost *A1* gene expression; both copies were hypermethylated in their promoter region and showed paramutagenic behavior that led to a permanent or temporary inactivation of *A1* gene expression in heterozgyotes with other *A1* transgene alleles. This inactivation correlated with hypermethylation of both alleles. The extent of paramutation was variable during plant development. This example then provides one mechanism for trans-inactivation: methylation provoked by a similarly methylated paramutagenic allele.

Another discovery that I believe is very relevant to the topic under review came from molecular analysis of a series of alleles at the *niv* locus in Antirrhinum that controls flower pigment production. The series includes closely related semidominant and recessive alleles where the semidominant alleles inactivate wild-type alleles in heterozygotes. Each of the semidominant alleles investigated has inversions and multiple copies of *niv* gene sequences. After consideration of several

models for how alleles with these sequence aberrations might confer dominance Coen *et al.* (31, 32) concluded that the inactivation of transcription probably was a consequence of a physical interaction between the semidominant and wild-type alleles.

The similarities between paramutation, the behavior of semidominant alleles, and the trans-inactivation phenomena seen with transgenes raise the question whether all the phenomena are consequences of the same collection of mechanisms. This reviewer believes they are.

Hypotheses to Explain the Trans-Inactivation Phenomena

Four categories of explanation have been offered by others to account for the phenomena (5–11, 13, 15, 30, 33–37). They are not mutually exclusive and no one is applicable to all the examples surveyed above. The first hypothesis suggests that the genes involved adopt an epigenetic state that affects gene expression after their physical interaction. In the second hypothesis gene expression is inhibited due to competition between genes for nondiffusible factors, essential for ordered transcription or translation, such as the nuclear matrix or nuclear envelope. The third and fourth hypotheses apply to cases in which transcription is not inhibited but specific mRNA degradation is the cause of loss of gene expression. The hypothesis debated most involves the production of unintended antisense RNA formation and the degradation of mRNA sense–antisense duplexes. The other involves the accumulation of higher levels of a specific RNA due to the addition of extra copies of its gene and the consequential degradation of all of this mRNA species by some unknown mechanism.

Below I outline nuclear processes that I believe provide a useful background to consider these and other explanations for the whole range of trans-inactivation phenomena reviewed above. This outline is then followed by a review of some of the supporting evidence for its constituent elements.

The processes to be considered first in summary and in more detail later are shown in Fig. 1. Current views of chromatin behavior and transcription imply that DNA in condensed chromatin is recruited into a decondensed form and becomes attached to the nuclear matrix, and regulated transcription is initiated. All these processes are programmed and regulated by complex interactions between a large array of regulatory proteins and DNA in chromatin.

If it is hypothesized that under certain conditions homologous DNA sequences interact in somatic nuclei to form a hybrid DNA duplex or triplex, then there could

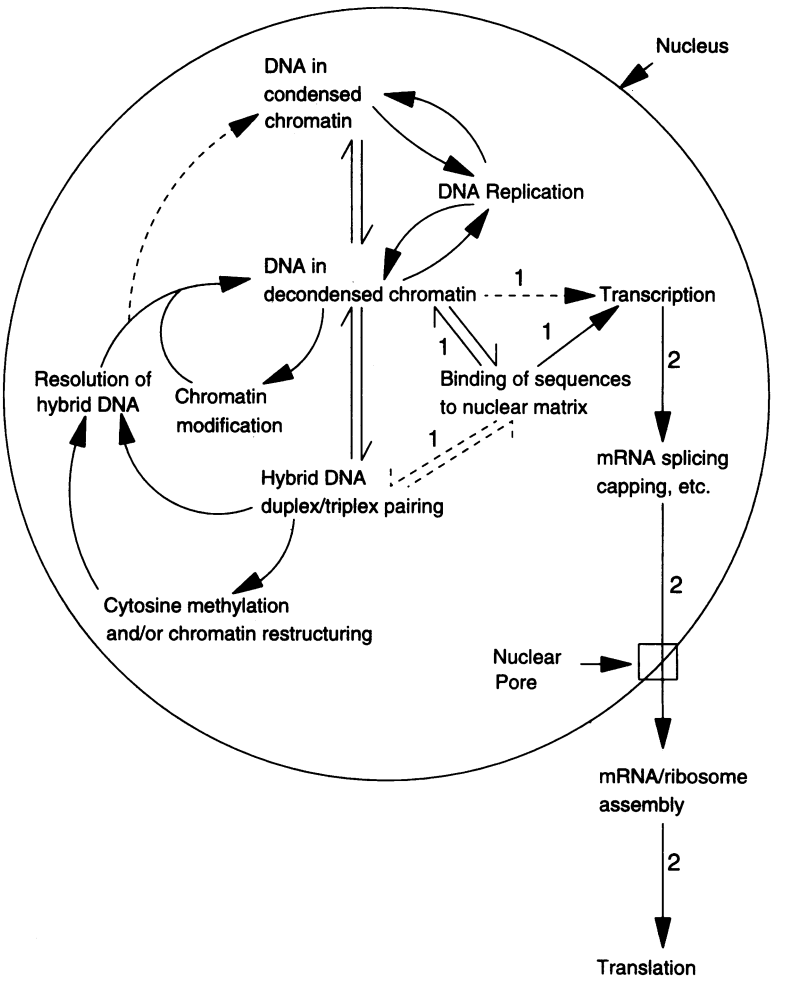

FIG. 1. Schematic illustrating the processes that could suffer aberrations to cause cosuppression and trans-inactivation. Details are described in the text.

be several unusual consequences. First, a new chromatin/DNA structure would result while the hybrid DNA was sustained. Second, there could be an exchange of chromatin proteins to create a new state of the chromatin/genes involved in the duplex. Third, if heteroduplex DNA were established, then different patterns of cytosine methylation could be imprinted into the participating DNA strands. These events should be viewed as being part of a dynamic reiterative cycle, as shown in Fig. 1, where reactions in segments of the cycle can occur many times per cell cycle or only spasmodically in different cells during development. On each occasion, the same or different copies of a sequence could be involved. The outcome of each stage could be different on each occasion, creating instability of allele chromatin structure and variation between alleles or, alternatively, the outcome could be constant, leaving allele chromatin structure very stable and homogeneous.

Any new chromatin state of a gene created via hybrid association or by gain of a protein or methylated cytosines via other routes could influence the subsequent chromatin condensation pattern as depicted in Fig. 1, with some states causing the gene to remain in condensed chromatin and silent. Hybrid DNA structures or those of the new chromatin states might interfere with binding of the genes to the nuclear matrix or with subsequent transcription to result in aberrant expression of some or all copies of the genes. The trans-inactivation would therefore result from interference in steps labeled 1 in Fig. 1.

When trans-inactivation is posttranscriptional, two routes can be hypothesized, affecting steps labeled 2 in Fig. 1. It can be envisaged that altered allelic chromatin states are transcribed, but in the wrong segment of the nucleus, and the mRNA–protein complexes are improperly or inefficiently transported through the nucleus, processed, and exported through the nuclear pore. They are consequently degraded. In an elaboration of this model, one could envisage that if mRNA transport processing, export, and possibly translation were inefficient, or excess mRNA accumulated due to aberrantly high levels of transcription, then an RNA-dependent RNA polymerase could synthesize antisense RNA

molecules on the mRNA templates. The antisense RNAs would form complexes with all the homologous sense mRNAs in the cell, and the complexes would be degraded before or after export from the nucleus. Alternatively, the RNA–RNA duplexes could be formed in the cytoplasm and prevent translation. The involvement of an RNA-dependent RNA polymerase has been recently postulated by Lindbo *et al.* (38) to explain the loss of transgene mRNA and viral RNA in virus-resistant tobacco plants containing a nuclear transgene encoding tobacco etch virus coat protein and infected with tobacco etch virus.

In the second route, it is envisaged that transgenes become incorporated into chromosomes under the control of promoters that generate antisense RNA to the transgene. The antisense RNA would inhibit RNA survival and translation as described above.

These scenarios could explain changes in the state of certain homologous genes, their imprinting and the stability/instability of new states, failure to be transcribed, or turnover of their mRNA.

The hypotheses have been divided into those that affect transcription and those that are posttranscriptional. These can be linked if a feedback system exists such that an accumulation of primary transcripts or antisense RNAs in the nucleus influences the chromatin or methylation state of a gene to affect its ability to efficiently participate in transcription. If this is the case, then inherited trans-inactivation of genes with sequence homology could occur due to excess on RNA buildup without hybrid DNA formation. The cycles of modification of chromatin structure shown in Fig. 1 without hybrid DNA formation are depicted to include this sort of possibility.

Evidence for the principal steps envisaged in these hypotheses is reviewed next.

(*i*) Interactions Between Loci with Homologous DNA Sequences. For interactions between loci with homologous sequences to occur, the DNA must be accessible. This suggests that the chromatin must be decondensed and DNA strands of the interacting loci be opened up—presumably during transcription, regulatory protein binding, replication, or recombination. This might explain the need for a promoter for trans-inactivation/cosuppression in some cases. Certain DNA structures such as the presence of inverted repeats might provoke alternative forms of genes that facilitate accessibility of DNA (32). Sequence-specific interaction implies hybrid chromatin or DNA duplex or triplex (39) formation, at least transiently, but this would need to be very efficient to account for the frequency of trans-inactivation observed, unless the chromatin/DNA that interacted was heritably imprinted. There is little evidence, direct or indirect, for such regular interactions involving all or most accessible chromatin sectors in somatic plant cells.

In yeast, such a DNA homology-searching process has been inferred to account for the equivalent frequencies of allelic and ectopic (in different position in the chromosome) meiotic recombination between homologous sequences (40). In Neurospora, duplicated sequences are detected by a homology searching/sensing process in the haploid nuclei of dikaryons before meiosis (41). The duplicated sequences are modified by cytosine methylation and a high proportion of the modified cytosines are substituted by thymidine. This repeat-induced point mutation process leads to destruction of gene function and elimination of such mutant progeny from the population. The DNA pairing process is not dependent on the meiotic chromosome pairing mechanisms (42). In *Ascobolus immersus* artificial gene repeats are also heritably inactivated, premeiotically, by cytosine methylation but other mutations are not introduced (43). DNA sequences can undergo several rounds of pairing so that multiple copies of a sequence can be inactivated sequentially (44). Thus, the pairing machinery does not distinguish between methylated and nonmethylated copies.

These observations on fungi provide a useful basis for considering possible mechanisms of trans-inactivation/cosuppression in plants. They provide a precedent for a process to control the expression of unusual duplicated sequences involving an efficient homology searching/sensing system. In these fungi, hybrid or paired DNA is presumably recognized in premeiotic cells, directly or indirectly, by a *de novo* methylase or a maintenance methylase that operates on hemimethylated DNA. Any parallel in plants would have to occur, presumably, in any somatic cell.

If duplicated sequences involving a transgene recognize each other and form hybrid DNA, even only transiently, how could different states of the transgenes or transgene and endogenous genes emerge and be inherited? There are two sorts of possibilities for this, with precedents in other kinds of organisms: restructuring of chromatin and DNA sequence modification. These processes are reviewed below.

(*ii*) Chromatin Restructuring. The case for modification of interacting loci via chromatin changes is based on observations established in *Drosophila*. Eleven examples have been reviewed recently (45) in which the expression of a gene is influenced by "sensing" the presence of another specific gene after some kind of localized somatic chromosome pairing (46–48). A transcription factor associated with the chromatin of one of the genes is postulated to interact also with the promoter/enhancer of another chromosome and modify the expression of this second locus (49). However, another model involving transacting regulatory RNAs as mediators of the effects has also been put forward (45). In a second type of interaction, a gene becomes inactivated by assuming a heterochromatic, repressing chromatin structure from its neighboring sequences. It then pairs with a wild-type allele on another chromosome and a mutant phenotype results because the chromatin structure of the wild-type gene is also converted to an inactive, heterochromatic form. This model is based on the ability of chromatin proteins determining heterochromatic condensation to initiate the condensation process, which then proceeds along the chromosome until some interfering components are encountered. When such initiating proteins are transferred to the wild-type gene via close pairing of chromosomes, the heterochromatic structure is imposed upon it also. Such modes of heterochromatic chromatin assembly are dominant to those normally determining the chromatin structure at the wild-type gene locus. When somatic pairing is disrupted in these cases, more normal gene expression ensues. In other cases, mutant phenotypes are enhanced when pairing is disrupted. The same is true for the trans-sensing effects on polytene chromosome puffs. In such cases, a mutant site will puff and accumulate mRNA when paired with its wild-type homologue but not when these chromosomes are desynapsed or remain homozygously paired.

An important issue for cosuppression and trans-inactivation in plants, as noted above, is how new states of the loci involved, however they are created, are stabilized and inherited. Other genetic studies on *Drosophila* have revealed gene products that appear to influence the state of gene activity through organismal development. Paro (50) has described a model that represents a way to imprint on/off transcription status into the higher-order chromatin structure surrounding a gene. He envisages chromatin of early embryos in a neutral state, allowing signals to activate specific gene loci. When the signals are not received, specific proteins interact with cis-regulatory elements of a gene and act as nucleation signals for a kind of heterochromatization, a local state that is clonally inherited through development. A parallel mechanism is envisaged whereby specific proteins would bind to specific loci to keep them open and available to developmentally controlled specific transcription factors.

The programming of transcriptional competence based on organization of chromatin structure is now receiving much attention in yeast and *Drosophila*, following the discovery of numerous

gene products that influence the expression of many genes and that are likely to work via influencing chromatin protein complexes (50–52). In summary, there is growing evidence from *Drosophila* and yeast, where the most detailed molecular genetic studies have been carried out that, in certain examples at least, chromatin states can be transferred between homologously paired chromatin segments, that competence for transcription is determined by chromatin structure, and that some chromatin states can be clonally inherited when not disturbed.

***(iii)* DNA Sequence Modification by Methylation.** The second mechanism that can be envisaged to account for the heritable modifications of the duplicated DNA sequences involved in cosuppression and trans-inactivation is methylation of different cytosine residues. This process could occur either in the hybrid DNA or after establishment of a chromatin structure by any route.

It is attractive to consider this process for many reasons: (*i*) the process occurs in all plant nuclei; (*ii*) cytosine methylation changes have been observed to correlate with at least some cases of cosuppression or trans-inactivation (e.g., see refs. 16, 23, and 30); (*iii*) it results in imprinting of DNA, and imprinted patterns are inherited due to the addition of methyl groups to new DNA strands based on the pattern in the old strand (53); (*iv*) enhanced cytosine methylation at key sites is known to correlate with modified chromatin structure and gene expression (54); and (*v*) it is known to be part of the mechanism used to silence artificially duplicated genes in fungi (41–43).

If hybrid DNA is formed between sequences of duplicate transgenes or between transgenes and endogenous sequences, then the DNA strands will be hemimethylated at many sites because the cytosine methylation pattern will differ between the two parental sequences. This is because when transgenes are introduced from *Agrobacterium* they are unmethylated and without a plant-determined chromatin structure. The process of cytosine methylation in CpG and CpXpG motifs must involve *de novo* methylation. It is likely that the stabilized pattern will reflect the chromosomal environment where it is inserted and random processes (30). The pattern may take many cell generations to stabilize. For these reasons, different insertions of the same transgene are likely to have different methylation patterns.

There is a very active methylase that recognizes hemimethylated DNA in plants (55) since 80% of the CpG and of the CpXpG sites are methylated and these have to be methylated after every round of replication. Therefore, given hybrid DNA and accessibility of these sequences to the appropriate methylase, the number of methylated sites in the hybrid DNA is likely to increase. Upon separation of the hybrid DNA strands and their assimilation back into their parent duplexes, hemimethylated parent templates would exist and would be methylated. Thus, after the interaction, both parental genes would be modified in the regions that formed hybrid DNA—generally in the direction of increased cytosine methylation. If the altered methylation pattern affected transcription directly or indirectly by the altered binding of regulatory proteins or by affecting the condensation pattern into a different chromatin conformation, then new heritable states of gene expression would have been created. The paired DNA sequences would have similar changes imprinted and thus might share similar changes in gene expression in many but not all cases. The process is usefully seen as a dynamic cyclical pathway (see Fig. 1) and multiple rounds of the cycle could occur in each cell cycle. To attain a stable state for all copies of a sequence, many cycles would probably be required.

The observed outcome would depend on the original state(s) of the transgene(s), and of the endogenous loci, and the role of specific cytosine residues in controlling levels of transcription, protein binding, and chromatin structure. Thus, gene expression could be reduced, increased, or unchanged.

A prediction of methylation of hemimethylated hybrid DNA is that homologous sequences should accumulate the same cytosine methylation pattern in the same cell lineage. This is testable. However, the pattern of methylated cytosines established could differ following the resolution of hybrid DNA duplexes in different cells, and thus chimeras would be produced, as has been observed (10, 11). The requirement for a transgene to provoke easily recognized cosuppression frequently may be because only newly inserted genes are in different, unregulated states of cytosine methylation and are therefore capable of giving rise to a variation in cytosine methylation and chromatin patterns via hybrid DNA formation.

***(iv)* Inhibition of mRNA Processing, Transport, Export, or Translation.** The first product of RNA transcription undergoes a series of complex processing steps including removal of intervening sequences, if present, capping at the 5′ end, and polyadenylylation at the 3′ end to produce the RNA that is transported through the nuclear pore for translation. How these processes occur within the nuclear architecture is not well documented. It has been proposed that RNAs are transported from the site of synthesis to the pore across a solid nuclear substructure distinct from chromatin (56). Evidence has been obtained that newly synthesized RNAs are associated with the nuclear matrix and pass along defined paths (refs. 57 and 58 but see ref. 59). The complexes that effect splicing may be localized in specific foci so RNAs requiring splicing may pass through such foci (60, 61). Association of mRNA with a splicing complex inhibits export; export is dependent on release from a spliceosome, and data have been reported to show that 5′ capping and a correct 3′ end are important for export (62).

Knowledge of these mechanisms implies that, for correct gene expression, RNA transcription may need to take place in an appropriate place in the nucleus and RNAs might traverse specific domains to be properly processed for export. If these requirements are not satisfied due to the relocalization of genes after their interaction and modification, then there could be delays in RNA transport and processing that result in degradation of the RNA.

Alternatively, if inefficient processing and transport were to occur, antisense RNA could be generated by using the accumulated RNA as template, because RNA-dependent RNA polymerase occurs in plant cells (63, 64). Its role and cellular location are unknown. The activity is induced after infection with viruses and after wounding. It is capable of making antisense RNA fragments from plant RNA and appears to display little template specificity. This activity, if nuclear, could synthesize antisense RNA in the nucleus. Alternatively, if cytoplasmic, it could synthesize antisense RNAs on exported RNAs present in excess in relation to regulated translation capacity or on RNAs inefficiently translated due to structural defects through aberrant processing. The antisense RNA products could remain cytoplasmic and prevent translation or enter the nucleus to interfere with RNA maturation.

The inactivation of sense mRNA to the transgene by unintended antisense RNA as an explanation for cosuppression has been debated by several authors (10, 11, 33–35). The authors envisaged that the antisense RNA would originate either from an active promoter in the host chromosome initiating transcription in the opposite direction to the transgene or from another gene promoter in the opposite orientation on the inserted T-DNA.

Efficient production of antisense RNA from a chromosome promoter would lead to down-regulation of sequence-specific gene expression and some antisense RNA has been detected in cosuppressed petunia plants (6, 7). However, there are specific pieces of evidence in examples of cosuppression and trans-inactivation that argue strongly against a chromosomal promoter being the source of antisense RNA and the sole component of a model

to explain all the phenomena associated with loss of gene expression. The evidence has been reviewed by others (10, 11, 34).

What is needed to evaluate the applicability of antisense RNA formation to the cause of down-regulation of homologous gene expression, in at least some examples of trans-inactivation by transgenes, is a much better understanding of how antisense RNA effects down-regulation of gene expression, measurements of antisense and sense RNA levels in the relevant cells and their nuclei before as well as after RNA degradation, and knowledge of the role of RNA-dependent RNA polymerase and of the ability of accumulated RNA products to feed back and interfere with transcription. It will also be important to discover the relationship between the mRNA turnover revealed by transgenes and endogenous posttranscriptional control systems that regulate mRNA turnover.

Implications of the Hypotheses

I have reviewed how chromatin interactions, nuclear control systems of chromatin structure, and DNA methylation mechanisms might produce changes in state in duplicated loci that interact physically or contain homologous sequences. The two key processes of chromatin and DNA modifications are not mutually exclusive but are inextricably linked. It is known that the distribution of methylated cytosine can affect the complement of chromatin proteins, chromatin condensation, and accessibility of the chromatin to transcription complexes and specific transcription factors. Similarly, the complement of chromatin proteins can influence the accessibility of DNA to methylase and hence the cytosine methylation pattern.

The occurrence and outcome of these potential processes in a transgenic plant would depend on the following: (*i*) whether two homologous sequences are able to pair (this might depend on their physical positions in the nucleus and on whether the DNA is accessible to allow homology-based recognition); (*ii*) the kinds of proteins bound to the loci and the nature of the nuclear processes that could modify the proteins at the loci; and (*iii*) the influence of the new chromatin state on its competence to influence chromatin binding to the nuclear matrix, active transcription, and efficient mRNA processing and export from the nucleus.

Different transgenes, or the same transgenes at different locations, might be expected to differ in *i* and *ii*, and hence *iii*, especially as the molecular environment of a gene is known to influence its adopted chromatin structure. If there is no significant variation in *ii* or no effect on *iii* then cosuppression or trans-inactivations would not be recognized. Failure to observe effects would not, however, necessarily imply that the cycles depicted in Fig. 1 do not occur. Nevertheless, it is important to ask whether the proposed cycles occur regularly to affect genes in all plant cells or whether they are peculiar to transgenes. It is extremely unlikely that the cellular processes are invoked only by transgenes—indeed, the examples to support the model include paramutation and the dominant alleles of endogenous plant genes reviewed earlier. However, it appears probable to me that transgenes adopt positions and chromatin structure that increase the probability that they participate in hybrid DNA formation. If they lack close-by sequences (scaffold attachment regions; SARs) to aid integration into the nuclear matrix, have strong constitutive promoters, and sustain an open chromatin structure, they are more likely than other genes to become involved in homology-searching processes, at least until they are inactivated into a closed, more silent chromatin structure.

Recent results have suggested that if matrix binding or flanking sequences that might contain a SAR are added to transgenes, then activity is much less sensitive to position effects (65). These important observations suggest that SARs may either interfere with hybrid DNA formation by enclosing DNA more efficiently or initiate a chromatin structure that is dominant and not significantly modified irregularly after hybrid DNA formation, or both. If the former, then hybrid DNA formation may be a default pathway of genes not properly associated with the nuclear matrix. I suspect that adapted genes, those evolved as part of the normal genome, are likely to have a much reduced probability of participating in hybrid DNA formation with their alleles or other members of their family or have adopted chromatin and DNA sequence environments to produce an outcome that is neutral with respect to gene expression. Unusual gene structures such as inverted duplication in complex transgenes or in mutants could provoke hybrid DNA pairing or its aberrant resolution.

Any model to explain trans-inactivations must account for how gene reactivation occurs following segregation of the duplicated loci into separate progeny. If the inactive state is dependent on continual and repeated interactions (hybrid DNA formation), then failure to form a hybrid DNA could rapidly lead to reversal. However, when the inactivation is due to an inherited state of methylation and chromatin configurations that is not reedited at meiosis, then the inactivated state might linger until new rounds of chromatin resetting had occurred. Cases of both types probably occur.

It is now essential to test the various hypotheses further. This will require studying the localization of genes and transcripts in nuclei, investigating chromatin structure and cytosine methylation patterns, searching for antisense RNAs, and exploring the activity of RNA-dependent RNA polymerase. It is also important to find out why transgenes in some locations do not involve cosuppression or trans-inactivation and find ways to control the phenomena. Uncontrolled inactivation of genes is a serious problem for the exploitation of plants in commerce, if not in the laboratory. On the other hand, inactivation of endogenous genes by insertion of homologous sequences into the genome opens up the opportunity to create mutants and identify the function of unknown sequences.

Significance of Cosuppression for Gene Organization and Genome Evolution

Whatever the mechanism(s) of cosuppression and trans-inactivation, the phenomena are likely to have played a substantial part in the evolution of genes, genomes, and mechanisms controlling gene expression. The immediate silencing of duplicated genes is a powerful event and is likely to have evolved for good reasons. This, in plants, is likely to have been to silence transposable elements and other repeats that can accumulate in the genome by mechanisms that make it difficult for them to be eliminated without very strong selection forces. Active transposable elements and retrotransposons are exceedingly mutagenic, and there would be strong selection in favor of mechanisms that can silence them immediately. New, active copies of transposable elements are in nonallelic positions with respect to other copies of homologous elements, similar to transgenes, and are likely to have new methylation and chromatin properties upon insertion and to be in accessible chromatin because of their mode of integration/movement in the chromosome.

The genes that are a regular part of the primary gene catalog, including members of multigene families, must have evolved mechanisms that either prevent them interacting or have evolved chromatin structures and mRNA processing and export strategies that upon interaction with the cellular processes outlined display a neutral outcome with respect to gene expression. This is consistent with genes evolving in particular environments of SARs and sequences that control chromatin structures, methylation patterns, nuclear locations, and adaptive levels of gene expression during plant development.

I wish to thank Rich Jorgensen, Marjorie Matzke, and Peter Meyer for teaching me so

much about the phenomena described in this review and for their comments on the manuscript.

1. Jones, J. D. G., Gilbert, D. E., Grady, K. L. & Jorgensen, R. A. (1987) *Mol. Gen. Genet.* **207**, 478–485.
2. Matzke, M. A., Primig, M., Trnovsky, J. & Matzke, A. J. M. (1989) *EMBO J.* **8**, 643–649.
3. Napoli, C., Lemieux, C. & Jorgensen, R. (1990) *Plant Cell* **2**, 279–289.
4. van der Krol, A. R., Mur, L. A., Bled, M., Mol, J. N. M. & Stuitje, A. R. (1990) *Plant Cell* **2**, 291–299.
5. Jorgensen, R. (1990) *Trends Biotechnol.* **8**, 340–344.
6. Mol, J., van Blokland, R. & Kooter, J. (1991) *Trends Biotechnol.* **9**, 182–183.
7. Kooter, J. M. & Mol, J. N. M. (1993) *Curr. Opin. Biotechnol.* **4**, 166–171.
8. Jorgensen, R. A. (1992) *AgBiotech News Info.* **4**, 265N–273N.
9. Jorgensen, R. A. (1993) *Philos. Trans. R. Soc. London B* **339**, 173–181.
10. Matzke, M. A. & Matzke, A. J. M. (1993) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 53–76.
11. Matzke, M. A., Matzke, A. J. M. & Schied, O. M. (1993) *Homologous Recombination and Gene Silencing in Plants*, ed. Paszkowski, J. (Kluwer, Dordrecht, The Netherlands).
12. Smith, C. J. S., Watson, C. F., Bird, C. R., Ray, J., Schuch, W. & Grierson, D. (1990) *Mol. Gen. Genet.* **224**, 477–481.
13. Assaad, F. F., Tucker, K. L. & Signer, E. R. (1993) *Plant Mol. Biol.* **22**, 1067–1085.
14. Scheid, O. M., Paszkowski, J. & Potrykus, I. (1991) *Mol. Gen. Genet.* **228**, 104–112.
15. Goring, D. R., Thomson, L. & Rothstein, S. J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1770–1774.
16. Hobbs, S. L. A., Warkentin, T. D. & DeLong, C. M. O. (1993) *Plant Mol. Biol.* **21**, 17–26.
17. Vaucheret, H. (1993) *C.R. Acad. Sci. Paris Life Sci.* **316**, 1471–1483.
18. Hobbs, S. L. A., Kpodar, P. & DeLong, C. M. O. (1990) *Plant Mol. Biol.* **15**, 851–864.
19. Peach, C. & Velten, J. (1991) *Plant Mol. Biol.* **17**, 49–60.
20. Linn, F., Heidmann, I., Saedler, H. & Meyer, P. (1990) *Mol. Gen. Genet.* **222**, 329–336.
21. Elkind, Y., Edwards, R., Mavandad, M., Hedrick, S. A., Ribak, O., Dixon, R. A. & Lamb, C. J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9057–9061.
22. Matzke, M. A., Neuhuber, F. & Matzke, A. J. M. (1993) *Mol. Gen. Genet.* **236**, 379–386.
23. Matzke, M. A. & Matzke, A. J. M. (1991) *Plant Mol. Biol.* **16**, 821–830.
24. Matzke, M. A. & Matzke, A. J. M. (1990) *Dev. Genet.* **11**, 214–223.
25. Matzke, A. J. M. & Matzke, M. A. (1993) *Semin. Dev. Biol.* **4**, 83–89.
26. de Carvalho, F., Gheysen, G., Kushnir, S., Van Montagu, M., Inze, D. & Castresana, C. (1992) *EMBO J.* **11**, 2592–2602.
27. Brink, R. A. (1973) *Annu. Rev. Genet.* **7**, 129–152.
28. Brink, R. A., Styles, E. D. & Axtell, J. D. (1968) *Science* **159**, 161–170.
29. Coe, E. H., Jr. (1966) *Genetics* **53**, 1935–1963.
30. Meyer, P., Heidmann, I. & Niedenhof, I. (1993) *Plant J.* **4**, 89–100.
31. Coen, E. S. & Carpenter, R. (1988) *EMBO J.* **7**, 877–883.
32. Bollmann, J., Carpenter, R. & Coen, E. S. (1991) *Plant Cell* **3**, 1327–1336.
33. Grierson, D., Fray, R. G., Hamilton, A. J., Smith, C. J. S. & Watson, C. F. (1991) *Trends Biotechnol.* **9**, 122–123.
34. Jorgensen, R. (1991) *Trends Biotechnol.* **9**, 266–267.
35. Mol, J., van Blokland, R. & Kooter, J. (1991) *Trends Biotechnol.* **9**, 182–183.
36. Hart, C. M., Fischer, B., Neuhaus, J. M. & Meins, F., Jr. (1992) *Mol. Gen. Genet.* **235**, 179–188.
37. Meins, R., Jr. (1989) *Annu. Rev. Genet.* **23**, 395–408.
38. Lindbo, J. A., Silva-Rosales, L., Proebsting, W. M. & Dougherty, W. G. (1993) *Plant Cell* **5**, 1749–1759.
39. Camerini-Otero, R. D. & Hsieh, P. (1993) *Cell* **73**, 217–223.
40. Haber, J. E., Leung, W.-Y., Borts, R. H. & Lichten, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1120–1124.
41. Selker, E. U. (1990) *Annu. Rev. Genet.* **24**, 579–613.
42. Foss, H. M. & Selker, E. U. (1991) *Mol. Gen. Genet.* **231**, 49–52.
43. Rhouanim, L., Rossignol, J.-L. & Faugeron, G. (1992) *EMBO J.* **11**, 4451–4457.
44. Faugeron, G., Rhounim, L. & Rossignol, J.-L. (1990) *Genetics* **124**, 585–591.
45. Tartof, K. D. & Henikoff, S. (1991) *Cell* **65**, 201–203.
46. Henikoff, S. (1990) *Curr. Opin. Genet. Dev.* **2**, 907–912.
47. Pirotta, V. (1990) *BioEssays* **12**, 409–414.
48. Wu, C.-T. (1993) *J. Cell Biol.* **120**, 587–590.
49. Dreesen, T. D., Henikoff, S. & Loughney, K. (1991) *Genes Dev.* **5**, 331–340.
50. Paro, R. (1990) *Trends Genet.* **6**, 416–421.
51. Travers, A. A. (1992) *Cell* **69**, 573–575.
52. Felsenfeld, F. (1992) *Nature (London)* **355**, 219–234.
53. Holliday, R. (1987) *Science* **238**, 163–170.
54. Selker, E. U. (1990) *Trends Biochem. Sci.* **15**, 103–107.
55. Kirnos, M. D., Aleksandrushkina, N. I. & Vanyushim, B. F. (1981) *Biokhimiya* **46**, 1458–1474.
56. Schroeder, H. C., Bachman, M., Diehl-Siefert, B. & Muller, W. (1987) *Prog. Nucleic Acid Res. Mol. Biol.* **34**, 89–142.
57. Lawrence, J. B., Singer, R. H. & Marselle, L. M. (1989) *Cell* **57**, 493–502.
58. Xing, Y. & Lawrence, J. B. (1991) *J. Cell Biol.* **112**, 1055–1063.
59. Rosbash, M. & Singer, R. H. (1993) *Cell* **75**, 399–401.
60. Mattaj, W. (1990) *Curr. Opin. Cell. Biol.* **2**, 528–538.
61. Maquet, L. E. (1991) *Curr. Opin. Cell. Biol.* **3**, 1004–1012.
62. Green, M. R. (1991) *Annu. Rev. Cell Biol.* **7**, 559–599.
63. Frankel-Conrat, H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 422–424.
64. Frankel-Conrat, H. (1986) *CRC Crit. Rev. Plant Sci.* **4**, 213–226.
65. Breyne, P., Van Montagu, M., Depicker, A. & Gheysen, G. (1992) *Plant Cell* **4**, 463–471.