

# Toward Monitoring Specific DNA Lesions in the Gene by Using Pollen Systems

by Michael Freeling\*

Specific gene systems expressed in cereal pollen could contribute uniquely to the problem of monitoring our environment for mutagens. This paper considers the development of a mutagen monitor with quantitative endpoints that reflect particular types of lesions at the DNA level, and lesions in particular components of the gene.

## Introduction

Two of the best understood genes in higher organisms, alcohol dehydrogenase-1 (*Adh1*) and waxy (*wx*), happen to be expressed in higher plant pollen grains. By using specific dysfunctional alleles of these genes, revertant frequencies may be quantified by pollen staining procedures, as evidenced by several papers in this volume. One could plant maize or barley lines homoallelic for any of hundreds of *wx* or *Adh1* alleles and quantify sundry insults to the DNA in terms of revertant grains (darkly stained) per total grains (light stained). In the discussion which follows, I hope to show that one of the unique contributions pollen systems may make as a mutagen monitor is to provide readily quantifiable endpoints which reflect particular types of lesions at the DNA level, or lesions in particular components of the gene. The Ames test (1) presently provides this sort of information for some transitions, transversions, and frameshifts in coding sequence DNA of a bacterial gene. However, the "gene" in higher organisms is certainly more complicated than its microbial analog. The abundance of repetitive DNA in the genomes of most higher organisms, including man, and the fact that many specific genes in higher organisms display coding sequences which are intervened by noncoding DNA should emphasize that fundamental differences exist between "a gene" in microbes and man. Pollen systems have the potential of monitoring genetic parameters unique to higher organisms. In

order to argue this assertion further, some background on current "theory of the gene" follows.

## Definition of the Gene and Its Functional Components

A "gene" is taken to be the most generous chromosomal unit of *cis* function. Therefore, a gene contains noncoding as well as coding sequences. Demerec's "cistron" or complementation group is one measure of the gene; Benzer's (2) use of the *cis-trans* test at the RII locus in T4 exemplifies this measure. Given that a gene may contain intervening sequences, *cis*-acting components beyond promoters, or even junk or spacer DNA, a superior means to measure the gene is by summing all chromosomal breakpoints which, by disconnecting the *cis* unit of function, either alter or obliterate the gene's product or alter the rate of cell/tissue/organ specificity, timing, fidelity, stability, circuitry with other genes, etc., of product expression. It might be useful to visualize the DNA of the gene as coding, regulatory component, spacer, recombinational, or junk.

It is obvious from comparisons of protein synthesis among tissues or organs of the same individual that gene expression is highly regulated both qualitatively (tissue-restricted gene expression, like  $\beta$ -globin) and quantitatively. In other words, genes are programmed to be on or off (high or low; early or late) in tissue-specific ways. In theory, this could be accomplished by utilizing tissue-specific DNA components of the gene, or by evoking a regulatory gene whose product affects the expression of certain genes but not others, or a combination of both

\* Department of Genetics, University of California, Berkeley, California 94720.

of these extremes. Variant and mutant alleles that specify regulatory-type changes in gene expression have been particularly valuable in identifying functional components of the gene.

The case for the existence of *cis*-acting quantitative gene components is built from phenomenological studies of mutant or variant alleles as compared to their wild-type progenitor alleles. Goldschmidt (3), in his review of the theory of the gene, argues aptly for a multicomponent gene integrated into a larger physiological system. McClintock (4) has studied numerous cases of derivatives from unstable alleles which are heritably altered in the timing or cellular pattern of expression (or mutation) during maize development. A quantitative series of *DS*-induced waxy alleles which differ in the ratio of branched to unbranched starch specified are exemplary (5). Studies using the *R* anthocyanine-conditioning locus in maize have identified an abundance of allelic variation involving quantitative parameters (6). A recent study by Kermicle (7) has mapped a quantitative component of *P* distal to the component that specifies organ-specificity. Unfortunately, chromosomal position effects (8) and early studies on unstable alleles suffer from lack of a biochemically accessible gene product. Specific alleles of genes in *Drosophila*, maize, and mice differ in *cis* regulatory properties involving tissue/organ specificity or developmental timing. Examples are esterase-5 (9) and alcohol dehydrogenase-1 in maize (10,12),  $\beta$ -glucuronidase (13),  $\beta$ -galactosidase (14), H2-antigen (15), aryl-sulfatase (16), and  $\beta$ -glucuronidase (17) in mice, and aldehyde oxidase (18) and alcohol dehydrogenase (19) in *Drosophila melanogaster*. The evidence from the above cases does not rigorously exclude the involvement of coding sequences. The behaviors of these alleles are clearly regulatory, are specified by a site in or near electrophoretically marked coding sequences and act in *cis*.

Chovnick and co-workers (20) have proved that a *cis*-acting site specifying the total amount of xanthine dehydrogenase per fly is genetically separable from the coding sequence.

One should not conclude from the above studies that all quantitative organ-specific gene regulation is a property of individual genes. For example, Abraham and Doane (21) found that the presence of  $\alpha$ -amylase in the midgut of *Drosophila melanogaster* is specified by a gene which acts on the  $\alpha$ -amylase structural gene in *trans*. Similarly, Dickinson (22) discovered three clear cases of specific negative gene regulation by soluble factors among closely related Hawaiian picture-winged *Drosophila*.

Taking all of these bits of information, it may be safely concluded that a gene in higher organisms is

more than a coding sequence and an on/off switch. Organ specific, quantitative components have been observed. As of yet, the intriguing behaviors of these variants or mutants have not been reduced to the level of nucleotide sequence and sequence arrangement.

## Gene Structure in Higher Organisms

A typical gene contains a structural gene component that carries information which specifies the amino acid sequence of its polypeptide product. Until about five years ago, it was assumed that the coding sequence was transcribed in one continuous 3' — 5' unit, called "the structural gene" component. Since 1977, it has become increasingly clear that many or most polypeptides in higher organisms are encoded in pieces of DNA with fragments of noncoding DNA between them (23-30). The structural gene is composed of coding and intervening sequences. As was shown clearly for a yeast tRNA gene (31), the intervening sequences are transcribed into a large, primary transcript that is subsequently processed within the nucleus by two or more cut-and-splices before a translatable message is pieced together and leaves the nucleus. The great majority of a structural gene's length can be intervening sequences. There is homology for nucleotides at splice-joints among widely divergent genes; this makes it possible that one or a few processing enzymes may be involved (32). From studies of  $\beta$ -globin gene fragments expressed in SV40, it seems that a pair of splice-joints is required for the production of stable messenger RNA (33). It is not clear whether or not the regulatory step of primary transcript — mRNA processing is used as an on/off switch, and/or for quantitative gene regulation. Several ongoing studies are seeking function and evolutionary origins of intervening sequences by comparing homogenous structural genes for conserved or diverged sequences or sequence locations (34).

It should be kept in mind that the structure of a gene need not be specified entirely by its nucleotide sequence. As with any other machine, the gene has various active and inactive three-dimensional shapes which could be controlled by various chromosomal RNAs or proteins, or by the gene's shape in the sister chromatid at the previous mitosis [see T. M. Sonnehorn (35) for concept of structural inheritance]. Wu, Wong, and Elgin (36) have used accessibility to chromatin digestion to show that *Drosophila* heat-shock genes change shape at about the same time they are induced to transcribe.

The 5' end of the message is translated first. For some primary transcripts, the 5' end either remains intact or a very few nucleotides are removed to form the 5' end of the message (37, 38). However, many messages have about 25% of their nucleotides in a 5' leader in front of start triplet ATG. Presumably, this later sequence is involved in translation by ribosomes (39). Similarly, the 3' end of the message is distal to the DNA stop triplet TAA: In addition to post-transcriptionally added poly(A) at the 3' terminus of many messages, there is also about a 20-nucleotide trailer which may contain an TA-rich conserved sequence (40).

About 30 nucleotides in the 3' direction of the start triplet ATG constitute an AT-rich presumptive promoter sequence which may be common to all eucaryotes (32, 41, 43). Although the primary structure of this "Pribnow box" is known, there is no causal evidence implying any function at all. Indeed, several intriguing stem-loop structures exist near coding sequences, but extrapolating from structure to function is not easily accomplished.

## Limitations of Reversion and Forward Mutation Assays

Mutant frequencies are generally quantified by using one of two general methods. The first method measures forward mutation, where a functional gene is mutated to dysfunction. The second method measures reversion, where a dysfunctional, mutant gene is reverted back to function. Although both methods measure DNA alteration, they provide vastly different endpoints.

From the previous summary on the nature of the gene, it should be clear that a forward mutation test with a dysfunction endpoint will register any mutant lesions which results in lowering the level of expression of a gene's product below the dysfunction threshold. For example, the waxy (*wx*) gene in cereals specifies a starch branching enzyme; branched starch in endosperm or pollen stains blue with an I<sub>2</sub>-KI solution. If *wx* does not function, then relatively unbranched starch stains red with I<sub>2</sub>-KI. It is certainly possible to stain nonmutant pollen with I<sub>2</sub>-KI and count red grains. In theory, such a forward mutation method would register as mutant any DNA lesion which leads to dysfunction: substitutions, deletions, or insertions in coding sequences; and any alteration in regulatory components which rendered the gene dysfunctional in pollen. Fortunately, pollen's requirement for a complete haploid genome acts as a filter for gross chromosomal aberrations. Unfortunately, it would be naive to expect every red pollen grain to be a *wx* mutant.

Perhaps they are merely sick. Stadler (43) estimated that the spontaneous *wx* mutant frequency in maize—scored as *wx* seeds—was below  $7 \times 10^{-7}$ . The frequency of red pollen in a normal plant is approximately  $10^{-5}$  (44). Therefore, these red gametophytes present an inscrutable genetic endpoint. With the *Adh1* system in maize pollen (45) a similar result is routine: up to 0.1% of healthy, *Adh1*<sup>+</sup> pollen does not stain positive (blue, opaque) for ADH. However, these unstained gametophytes proved to be inviable (Cheng and Freeling, unpublished). In units of ADH<sup>+</sup>/germinatable pollen, the *Adh1*<sup>+</sup> → ADH<sup>-</sup> spontaneous forward mutant frequency was estimated to be below  $2 \times 10^{-7}$  (46).

While forward mutant assays are unbiased in that the entire, naturally occurring gene is a target, obtaining data with genetic resolution at  $10^{-6}$  may not be possible. For the *Adh1* gene, allyl alcohol vapor treatment can select rare ADH<sup>-</sup> ungerminated gametophytes among millions of ADH<sup>+</sup>s. This could permit recovery and molecular characterization of the spectra of mutants generated by a treatment or environment (47). It should be noted that mutant spectra data are quantitative to the extent that a distribution of mutant types could be displayed, but reliable estimates of mutation rates can not be reasonably expected from mutant recovery frequencies after selection.

Revertant assays are easier to quantify reliably because the restoration of an enzyme activity should be unrelated to disease or inviability. The term "phenotypic revertant" is meant to include virtually any way a cell regains the function of a mutant gene. "True reversion" is where the change in DNA sequence which caused the mutant behavior is reversed to the original sequence. "Second-site" reversion includes all other alterations in the coding sequence which "complement" the original mutant lesion via some intrapolypeptide folding mechanism. Phenotypic reversion also includes ways to regain a gene's function without altering the gene at all: suppression by unique tRNAs (supersuppression) and other mechanisms, activation of a silent isoenzyme gene, conversion by an isoenzyme gene, and so on. In the absence of revertant recovery and characterization, or other indirect controls, phenotypic revertants are a complex endpoint.

In order to estimate the spontaneous phenotypic revertant frequency for specific genes in pollen, one should begin with several dysfunctional mutants. Using 14 independent alleles of *wx* in maize, Nelson (48) obtained a median revertant frequency (blue pollen/total pollen) of  $7 \times 10^{-6}$ . I used seven *Adh1*-S-deficient alleles induced with ethyl methanesulfonate by D. Schwartz to obtain a median revertant fre-

quency of  $5.7 \times 10^{-6}$  (46). These values are in good agreement and are well below typical intragenic recombination frequencies ( $2 \times 10^{-5}$ - $10^{-3}$ ). Even so, these five per million spontaneous revertant frequencies are much higher than expected on the basis of the below 0.5 per million estimates of spontaneous forward mutant frequencies for maize *wx* and *Adh1*. Such excessive revertant frequencies are not uncommon among eukaryotes; I have compiled the relevant data elsewhere (46). As a rule, it should be easier to wreck a machine than repair it. A solution to this apparent excessive reversion paradox is necessary in order to understand what revertants of an allele really meant at the DNA level.

When an allele generates a revertant frequency above  $10^{-5}$ , it is called "unstable" or "mutable." There is much evidence that mutable alleles reflect unstable integration of a piece of DNA in the gene (49,50). There is a tendency to pick unstable alleles for mutagen monitoring because fewer pollen grains need be scored in reversion assays. A *wx* allele in common use (44,51) has high spontaneous reversion frequency. While I can imagine the judicious use of mutable alleles as monitors, their present use is premature.

## The Future of Pollen Mutagen Monitors

If specific gene reversion systems in cereal pollen are to be used, they should contribute uniquely where other monitors fail. As plants, cereals are certainly appropriate when there is some risk that crops are activating chemicals which might enter the food chain. As inexpensive, *in situ* monitors for volatile mutagens, plant systems excel. A third unique contribution of pollen systems is to monitor specific lesions in DNA or lesions in specific gene components. In previous sections of this paper, I have shown that the "gene" in higher organisms is composed of quantitative organ-specific functional components and is structurally complex as well. Such complicated genes seem to be confined to organisms displaying the classical embryological phenomena of determination, competence and somatic heredity. Ideally, pollen systems would monitor mutational events which could not be monitored in the Ames or any other microbial test. If lesions of a particular type or in a particular gene component prove diagnostic for one or more cancers, the monitor would be most powerful.

I suggest two specific approaches toward developing truly elegant pollen mutagen monitors.

**Approach No. 1.** This approach requires com-

plete DNA level understanding of select mutant alleles with their normal progenitor alleles, it is possible to identify the exact change in DNA which causes the mutant behavior. This comparison could be done at the level of amino acid sequencing if the mutant site were in a coding sequence, but would require DNA-level comparisons if the site were in noncoding gene components. Particular alleles would then be chosen. Hypothetical examples include an insertion in coding sequence; a small inversion in nontranscribed "quantitative component;" a missing splice-joint; and so forth. A spectrum of revertants of these known lesions would be recovered and characterized molecularly to better understand just what a "phenotypic revertant" of a particular mutant allele means.

**Approach No. 2.** Comparative, quantitative mutagenesis using several molecularly distinct alleles and several known mutagens and carcinogens. Do active carcinogens generate certain kinds of DNA lesion (like chromosomal breaks) more efficiently than others (like point-mutations)? I have argued elsewhere (12) that regulatory-type mutants at a gene seem to result from chromosomal breakage-type events as opposed to small changes confined to coding sequences. If certain kinds of DNA lesions preferentially cause cancers, then we need a monitor for such diagnostic lesions. I believe that pollen systems—particularly *Adh1* in maize, but other systems to a high-resolution, mutagen monitor with DNA-level endpoints.

This review and analysis derives from experimental studies supported by Public Health Service, NIH GM21734 and by a contract with Lawrence Berkeley Laboratory.

## REFERENCES

- McCann, J., and Ames, B. N. Detection of carcinogens as mutagens in the *Salmonella*/microsome test: assay of 300 chemicals. *Proc. Natl. Acad. Sci. (U. S.)* 73: 950 (1976).
- Benzer, S. On the topology of the genetic fine structure. *Proc. Natl. Acad. Sci. (U. S.)* 45: 1607 (1959).
- Goldschmidt, R. B. Chromosomes and genes. *Cold Spring Harbor Symp. Quant. Biol.* 16: 1 (1951).
- McClintock, B. Chromosome organization and genic expression. *Cold Spring Harbor Symp. Quant. Biol.* 16: 13 (1951).
- McClintock, B. Mutable loci in maize. *Carnegie Inst. Wash. Yr. Bk.* 47: 155 (1948).
- Brink, R. A. Paramutation. *Ann. Rev. Genet.* 7: 129 (1973).
- Kermiele, J. L. Probing the component structure of a maize gene with transposable controlling elements. *Science* 208: 1457 (1980).
- Baker, W. K. Position effect variegation. *Adv. Genetics* 14: 133 (1968).
- Schwartz, D. Genetics studies on mutant enzymes in maize. III. Control of gene action in the synthesis of pH 7.5 esterase. *Genetics* 47: 1609 (1962).
- Schwartz, D. Genetic control of alcohol dehydrogenase—a competition model for regulation of gene action. *Genetics* 67: 411 (1971).

11. Freeling, M., and Cheng, D. S. K. Radiation-induced alcohol dehydrogenase mutants in maize following allyl alcohol selection of pollen. *Genet. Res.* 31: 107 (1978).
12. Freeling, M., and Woodman, J. C. Regulatory variant and mutant alleles in higher organisms and their possible origin via chromosomal breaks. In: *The Plant Seed*. I. Rubenstein, et al., Eds., Academic Press, New York, 1979, pp. 85-111.
13. Paigen, K. The genetic control of enzyme activity during differentiation. *Proc. Natl. Acad. Sci. (U. S. A.)* 47: 1641 (1961).
14. Paigen, K., Meister, J., Felton, J., and Chapman, V. Genetic determination of beta-galactosidase developmental program in mice. *Cell* 9: 533 (1976).
15. Boubelik, M., Lengoerová, A., Bailey, D. W. and Matoušek, V. A model for genetic analysis of programmed gene expression as reflected in the development of membrane antigens. *Develop. Biol.* 47: 206 (1975).
16. Daniel, W. L. Genetic control of heat sensitivity and activity level of murine arylsulfatase B. *Biochem. Genet.* 14: 1003 (1976).
17. Dofuku, R., Tettenborn, V., and Ohno, S. Further characterization of *O*<sup>s</sup> mutation of mouse  $\beta$ -glucuronidase locus. *Nature* 238: 216 (1971).
18. Dickenson, W. J. A genetic locus affecting the developmental expression of an enzyme in *Drosophila melanogaster*. *Develop. Biol.* 42: 131 (1975).
19. Dickenson, W. J., and Carson, H. L. Regulation of the tissue specificity of an enzyme by a cis-acting genetic element. *Proc. Natl. Acad. Sci. (U. S. A.)* 76: 4559 (1979).
20. Chovnick, A., Gelbart, W. and McCarron, M. Organization of the rosy locus in *Drosophila melanogaster*. *Cell* 11: 1 (1977).
21. Abraham, I., and Doane, W. W. Genetic regulation of tissue-specific expression of *Amylase* structural genes in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. (U. S. A.)* 75: 4446 (1978).
22. Dickenson, W. J. Regulation of the tissue specificity of enzymes by trans-acting factors in *Drosophila* hybrids. *Science* 207: 995 (1980).
23. Breathnach, R., Mandel, J. L., and Chambon, P. Ovalbumin gene is split in chicken DNA. *Nature* 70: 314 (1977).
24. Leder, P., Tilghman, S., Tiemeier, D. C., Polsky, F. T., Seidman, J. G., Edgell, M. H., Enquist, L. W., Leder, A., and Norman, B. The cloning of mouse globin and surrounding gene sequences in bacteriophage  $\lambda$ . *Cold Spring Harbor Symp. Quant. Biol.* 42: 915 (1977).
25. Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V., and Leder, P. Intervening sequence of DNA identified in the structural portion of a mouse  $\beta$ -globin gene. *Proc. Natl. Acad. Sci. (U. S. A.)* 75: 725 (1978).
26. Jeffreys, A. M., and Flavell, R. A. The rabbit  $\beta$ -globin gene contains a large insert in the coding sequence. *Cell* 12: 1097 (1977).
27. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., and Maniatis, T. The isolation and characterization of linked  $\delta$ - and  $\beta$ -globin genes from a cloned library of human DNA. *Cell* 15: 1157 (1978).
28. Engel, J. D., and Dodgson, J. B. Analysis of the adult and embryonic chicken globin genes in chromosomal DNA. *J. Biol. Chem.* 253: 8239 (1978).
29. Catterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanaka, Y., and Brownlee, G. G. Nucleotide sequence homology at 12 intron-exon junctions in the chick ovalbumin gene. *Nature* 275: 510 (1978).
30. Cordel, B., Bell, G., Tischer, E., DeNoto, F. M., Vilnich, A., Pictet, R., Rutter, W. J., and Goodman, H. M. Isolation and characterization of a cloned rat insulin gene. *Cell* 18: 533 (1979).
31. Knapp, G., Beckman, J. S., Johnson, P. F., Fuhrman, S. A., and Abelson, J. Transcription and processing of intervening sequences in yeast tRNA genes. *Cell* 14: 221 (1978).
32. Tsujimoto, Y., and Suzuki, Y. Structural analysis of the fibroin gene at the 5' end and its surrounding regions. *Cell* 16: 425 (1979).
33. Hamer, D. H., and Leder, P. SV40 recombinants carrying a functional RNA splice junction and polyadenylation site from the chromosomal mouse  $\beta^{maj}$  globin gene. *Cell* 17: 737 (1979).
34. Nishioka, Y., Leder, A., and Leder, P. Unusual  $\alpha$ -like globin gene that has clearly lost both globin intervening sequences. *Proc. Natl. Acad. Sci. (U. S. A.)* 77: 2806 (1980).
35. Sonneborn, T. M. Gene action in development. *Proc. Roy. Soc. (London)* 176: 347 (1970).
36. Wu, C., Wong, Y.-C., and Elgin, S. The chromatin structure of specific genes. II. Disruption of chromatin structure during gene activity. *Cell* 16: 87 (1979).
37. Curtis, P. J., Mantei, N., and Weissman, C. Characterization and kinetics of synthesis of 15S  $\beta$ -globin RNA, a putative precursor of  $\beta$ -globin mRNA. *Cold Spring Harbor Symp. Quant. Biol.* 42: 971 (1977).
38. Konkel, D. A., Tilghman, S. M., and Leder, P. The sequences of the chromosomal mouse  $\beta$ -globin major genes: homologies in capping, splicing and poly(A) sites. *Cell* 15: 1125 (1978).
39. Kozak, M. How do eukaryotic ribosomes select initiation regions in messenger RNA. *Cell* 15: 1109 (1978).
40. Proudfoot, N. M., and Brownlee, G. G. 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature* 263: 211 (1976).
41. Nishioka, Y., and Leder, P. The complete sequence of a chromosomal mouse  $\alpha$ -globin gene reveals elements conserved throughout vertebrate evolution. *Cell* 19: 875 (1979).
42. Gannon, F., O'Hare, K., Perrin, F., LePenec, J. P., Benoist, C., Cockett, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., and Chambon, P. Organization and sequences at the 5' end of a cloned complete ovalbumin gene. *Nature* 278: 428 (1979).
43. Stadler, L. J. Some observations on gene variability and spontaneous mutation. In: *The Spragg Memorial Lectures on Plant Breeding (Third Series)*, Michigan State College, East Lansing, 1942, pp. 3-15.
44. Plewa, M. J., and Wagner, E. D. Germinal cell mutagenesis in specially designed maize genotypes. *Environ. Health Perspect.* 37: 61 (1980).
45. Freeling, M. Maize *Adh1* as a monitor of environmental mutagens. *Environ. Health Perspect.* 27: 91 (1978).
46. Freeling, M. Spontaneous forward mutation versus reversion frequencies for maize *Adh1* in pollen. *Nature* 267: 154 (1977).
47. Schwartz, D. *Adh* locus in maize for detection of mutagens in the environment. *Environ. Health Perspect.* 37: 75 (1980).
48. Nelson, O. E. The waxy locus in maize. II. The location of the controlling element alleles. *Genetics* 60: 507 (1968).
49. McClintock, B. Genetic systems regulating gene expression during development. *Develop. Biol. (Suppl.)* 1: 84 (1967).
50. Green, M. M. A case for DNA insertion mutations in *Drosophila melanogaster*. In: *DNA Insertion Elements, Plasmids and Episomes*. A. Bikhark, J. Shapiro, and A. Adhya, Eds., Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1977, pp. 434-446.
51. Lower, W. Unpublished data.