3. Plant genes

Understanding mechanisms of gene expression

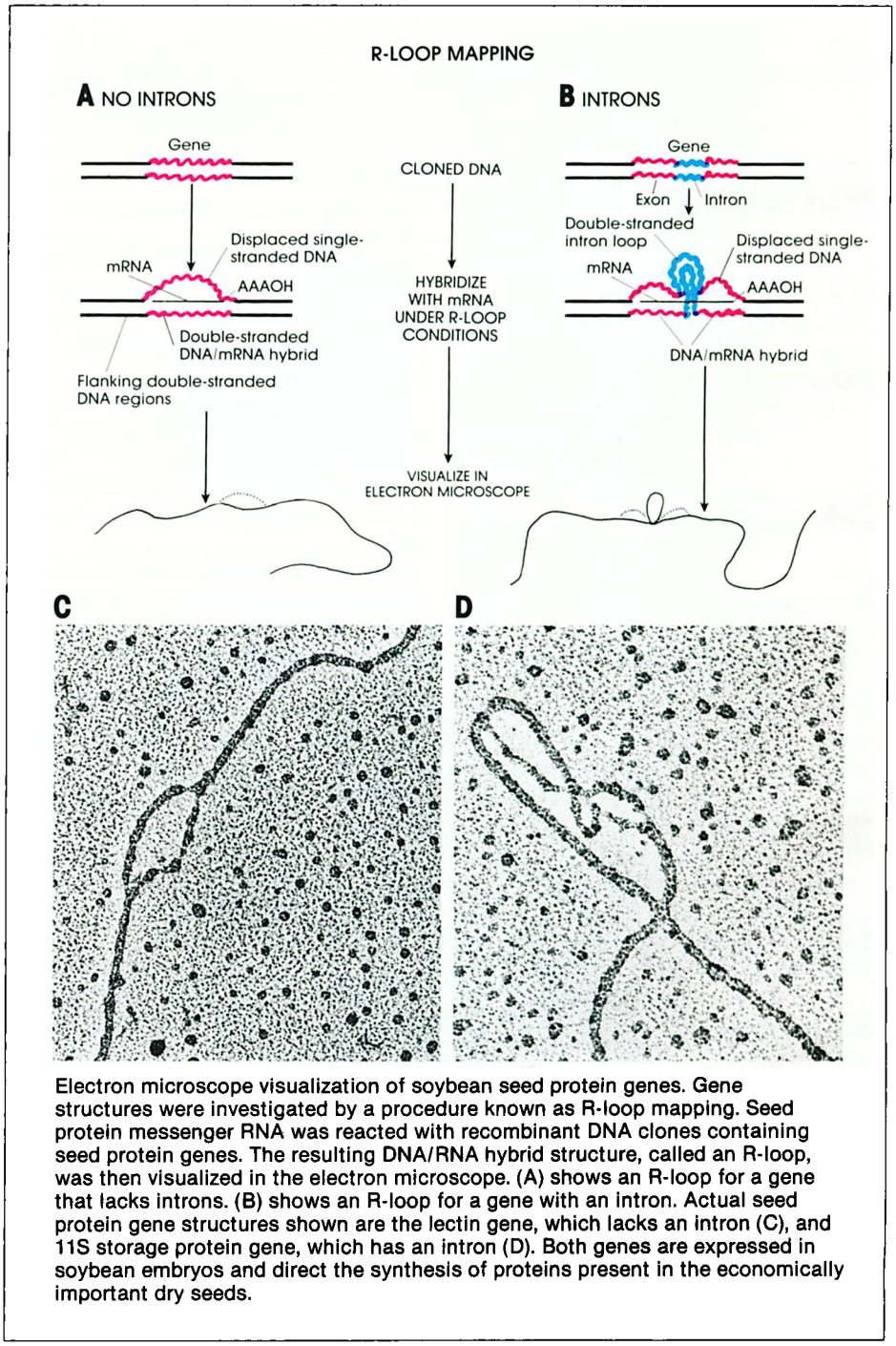
William E. Timberlake, Associate Professor, Bacteriology, Plant Growth Laboratory, Davis, and
Robert B. Goldberg, Associate Professor, Biology, Los Angeles

Underlying the developing technology of plant genetic engineering is the knowledge that specific traits result from expression of discrete regions of the DNA molecules present in each cell. These regions comprise the familiar, basic units of biological information and heredity, the genes. Proteins, the molecules whose activities finally determine the form and function of the plant, are encoded by the linear arrangement of purine and pyrimidine bases in the DNA of individual genes. In plant breeding, genes conferring desirable characteristics are transferred from one genetic background to another by "crossing." In contrast, the new approaches to plant improvement do not depend entirely on naturally occurring mechanisms for genetic recombination. Rather, specific fragments of DNA and, thus, specific genes, can be obtained from any organism, modified in the laboratory as desired, and, in theory, introduced into any plant.

The ultimate success of plant genetic engineering depends entirely on new information being made available by basic research in plant molecular biology. It is essential that we elucidate the structure and organization of important plant genes in order to understand the molecular processes that control their orderly expression. We must develop methods for identifying potentially useful plant genes at the molecular level, obtaining DNA fragments in which they are contained, and constructively altering them outside the plant. We need to determine the mechanisms that coordinate the expression of genes during plant growth and development. Only then can efficient procedures be designed for introducing novel genetic information into plants in a biologically relevant way.

Although much remains to be learned, and advances in plant molecular biology have not been as rapid as in bacterial, fungal, and animal molecular biology, a large body of information on plant gene structure, organization, and expression has begun to accumulate.

The molecular organization of plants is exceedingly complex. Their chromosomes, like



Electron microscope visualization of soybean seed protein genes. Gene structures were investigated by a procedure known as R-loop mapping. Seed protein messenger RNA was reacted with recombinant DNA clones containing seed protein genes. The resulting DNA/RNA hybrid structure, called an R-loop, was then visualized in the electron microscope. (A) shows an R-loop for a gene that lacks introns. (B) shows an R-loop for a gene with an intron. Actual seed protein gene structures shown are the lectin gene, which lacks an intron (C), and 11S storage protein gene, which has an intron (D). Both genes are expressed in soybean embryos and direct the synthesis of proteins present in the economically important dry seeds.

those of animals, contain enough genetic information to code for hundreds of thousands of different proteins. Most of the sequences present in the DNA exist in only single copies, and the vast majority of the genes are thought to occur in this type of DNA. However, a large proportion of plant DNA consists of sequences that are repeated from a few to several thousand times in each nucleus. Some of these reiterated DNA components—for example, those coding for certain specialized protein products and for ribosomal RNA—have known functions. The roles the others play in the life of the plant are currently only a matter for speculation.

Estimates have now been made of the actual proportion of a plant's genetic information that is transcribed into messenger RNA molecules, the nucleic acid sequences that directly specify the synthesis of proteins. Approximately 11 percent of the unique DNA, or 5 percent of the entire genome, of tobacco is expressed as messenger RNA in the sporophyte plant—equivalent to about 60,000 different genes needed to program and maintain this phase of the life cycle. Some of these genes are transcribed into messenger RNAs present in all of the organ systems of the plant. It is thought that these genes code for proteins whose activities are required for the lives of all cells, the "housekeeping" genes. Many others, however, are expressed in only one organ system. These genes presumably are responsible for the specialized functions of the cells comprising the leaves, roots, stems, petals, anthers, and ovaries.

The actual physiological functions of the vast majority of plant genes are simply unknown, even though they are certainly of substantial biological importance. The problem is that the cellular levels of most messenger RNAs are exceedingly low, making them difficult to identify and characterize. However, many of the genes coding for rare-class messenger RNAs are also expected to be of considerable agricultural significance. The question is, how can such genes be more readily isolated and studied?

Information is only now beginning to accumulate on the molecular mechanisms that control the pattern of gene expression in plant cells. The activities of genes may be regulated at any one of several biochemical levels. For example, the soybean, like other crop plants, has a set of genes that codes for seed storage proteins. These genes are transcribed into messenger RNAs that accumulate to high levels in developing seeds, but their expression cannot be detected elsewhere in the soybean plant, even by extremely sensitive assay methods. Such results argue that the developmentally specific expression of these genes is controlled at the level of transcription, as is typical of lower organisms. In contrast, many of the genes that encode messenger RNAs occurring specifically in one organ system are also transcriptionally active in other organ systems. The organ-specific expression of these genes is the result of post-transcriptional control mechanisms. Some RNAs present in the nucleus are selected for transport to the cytoplasm, the site of cellular protein synthesis; others are rapidly broken down within the nucleus itself.

We know virtually nothing about specific biochemical events that underly these regulatory processes. Such information is of seminal importance to genetic engineering. For example, transfer into a plant's chromosomes of a gene conferring a desirable alteration in leaf physiology could kill the plant if the gene were also expressed in the roots. We have to ensure that genes introduced into plants in new ways will be expressed in the appropriate tissue and at the proper time in development.

The study of plant gene expression has recently been facilitated by application of recombinant DNA technology. Entire plant genomes have been fragmented and introduced into bacterial viruses to form recombinant DNA sequence "libraries." Similarly, DNA copies of messenger RNA species have been joined to bacterial plasmids and propagated in bacteria. Individual recombinant DNA clones containing plant genes of interest have been isolated from these collections and analyzed, revealing several features of gene structure and organization that are clearly relevant to plant genetic engineering.

The structure of plant genes is variable. Some appear to be similar to genes typical of lower eukaryotic organisms, such as fungi, in that they comprise continuous protein coding regions—that is, they are colinear with their messenger RNA products. Most, however, are more complex, as is characteristic of genes in birds and mammals. In these cases, regions of the genes that code for protein, called exons, are interrupted by noncoding sequences, termed introns. The cell modifies the RNA products of such genes in a reaction whereby the noncoding regions are excised and the coding sequences spliced together to yield mature messenger RNAs that can direct the synthesis of functional proteins. It has been estimated that about 75 percent of a typical plant gene consists of one or more introns.

Recombinant DNA and DNA sequencing technologies have also allowed investigation of DNA regions adjacent to plant genes. Sequences required for the initiation of RNA synthesis, excision of introns, and post-transcriptional addition of the polyadenylic acid sequences normally found in messenger RNAs have been identified and are similar to those observed in other types of organisms. It has been discovered that plant genes having very different patterns of expression may be closely associated with one another in the genome. In contrast, genes with closely related forms and functions may be dispersed throughout the genome.

Although the biological significance of these features of gene structure and organization is not entirely clear, such observations are yielding valuable clues to the processes controlling plant gene expression and are suggesting directions for future research. They are also providing the basis for modifying plant genes in the laboratory, adapting genes from other organisms or plant species for use in solving specific agricultural problems, and constructing vectors to introduce novel genetic information into plants. The following reports present some examples of specific plant gene systems currently under intensive examination by U.C. research groups.