5d.

The Plant Cell, Vol. 1, 229–239, February 1989, © 1989 American Society of Plant Physiologists

Transcription, Organization, and Sequence of an Auxin-Regulated Gene Cluster in Soybean

Bruce A. McClure,[a,1] Gretchen Hagen,[a] Christopher S. Brown,[a] Melissa A. Gee,[b] and Tom J. Guilfoyle[a,2]

[a] Department of Biochemistry, University of Missouri-Columbia, Columbia, Missouri 65211
[b] Department of Biology, University of Missouri-Columbia, Columbia, Missouri 65211

We have characterized a soybean gene cluster that encodes a group of auxin-regulated RNAs (small auxin up RNAs). DNA sequencing of a portion of the locus reveals five homologous genes, spaced at intervals of about 1.25 kilobases and transcribed in alternate directions. At least three of the genes are transcriptionally regulated by auxin. An increase in the rate of transcription is detected 10 min after application of auxin to soybean elongating hypocotyl sections. Each of the genes contains an open reading frame that could encode a protein of 9 kilodaltons to 10.5 kilodaltons. Sequence comparisons among the five genes reveal several areas of high homology. Two regions of high homology begin about 250 base pairs upstream of the open reading frames and two regions of homology have been identified in sequences downstream of the open reading frames. One of the latter sequences occurs in the 3′-untranslated region of the RNAs. The other occurs far downstream, 618 base pairs to 741 base pairs from the stop codon. Conservation of these sequences among the five different genes suggests that they may be important for the regulation of expression of the genes.

INTRODUCTION

Auxins are a group of plant hormones that include natural compounds such as indole-3-acetic acid and synthetic compounds such as 2,4-dichlorophenoxyacetic acid (2,4-D). Auxins play a role in many plant developmental processes including cell elongation and cell division. The cell elongation response to auxin has been the most extensively studied, primarily due to the rapidity of the response (Evans, 1974, 1985). In excised soybean elongating hypocotyl sections (EHS), auxin-induced elongation may be observed after a lag of about 12 min (Vanderhoef, 1980). There has been some controversy concerning whether the events that occur during this lag period involve the induction of specific auxin-regulated genes. One of the main arguments against the gene induction hypothesis has been that the elongation response occurs too rapidly for the auxin-induced gene products to accumulate (Evans and Ray, 1969; Evans, 1974). This argument has been made less tenable with the observations that specific, auxin-induced changes in gene expression occur concomitant with (Walker and Key, 1982; Hagen et al., 1984; Theologis et al., 1985) or well before (McClure and Guilfoyle, 1987) the cell elongation response. In addition, inhibitor studies (for review see Key, 1969) also implicate gene expression in the cell elongation response. Thus, it seems likely that auxin-regulated gene expression plays some role in rapid plant growth responses such as cell elongation. However, the mechanism of auxin-induced gene activation remains to be determined.

Recently, auxin-regulated gene expression has been studied by several groups using cloned cDNAs from soybean hypocotyl (Baulcombe and Key, 1980; Walker and Key, 1982; Hagen et al., 1984; McClure and Guilfoyle, 1987), pea epicotyl (Theologis et al., 1985), and tobacco cell cultures (van der Zaal et al., 1987). These studies have shown that auxin rapidly and specifically alters the abundance of a few RNA species (for review see Guilfoyle, 1986; Theologis, 1986; Hagen, 1987). Using isolated soybean nuclei and in vitro nuclear run-on assays, it has been shown that at least some of these sequences are directly regulated at the level of transcription (Hagen and Guilfoyle, 1985). Genomic DNA sequences of a few of these auxin-regulated genes have been reported (Ainley et al., 1988; Czarnecka et al., 1988; Hagen et al., 1988).

Recently, we have characterized a group of small RNAs regulated by auxin in soybean (McClure and Guilfoyle, 1987), which we refer to as SAURs (small auxin up RNAs). In the EHS, accumulation of the SAURs is auxin-specific and begins within 2.5 min after treatment with 50 μM 2,4-D. This rapid response suggests that the RNA accumulation is tightly coupled to the auxin stimulus. Recently, we have shown that SAUR expression is dynamically regulated in vivo in gravistimulated soybean hypocotyls (Mc-

[1] Current address: Plant Cell Biology Research Center, University of Melbourne, School of Botany, Parkville, Victoria 3052, Australia.
[2] To whom correspondence should be addressed.

Clure and Guilfoyle, 1989). In this paper, we show that the genes encoding the SAURs are regulated at the level of transcription. We also show that the genes are tightly clustered in the soybean genome, and that the sequences of five SAUR genes contain several regions of high homology.

RESULTS

Transcriptional Activation

Accumulation of SAURs is under dynamic control in both excised (McClure and Guilfoyle, 1987) and intact (B. McClure and T. Guilfoyle, manuscript submitted) organs. Since we wish to study the accumulation of these RNAs as an auxin response, it is important to establish the level at which the response is regulated (i.e. transcriptional and/or posttranscriptional regulation). Figure 1 shows results of in vitro nuclear run-on experiments designed to probe the transcriptional regulation of SAUR accumulation. Gene-specific cDNAs corresponding to 3′-untranslated regions of three of the sequenced genes in the SAUR locus were used to test the transcriptional inducibility of these genes. The results in Figure 1A show that the rate of transcription of genes 6B, 10A5, and 15A is dramatically increased in the presence of auxin. Low levels of α-amanitin (0.5 μg/ml) strongly inhibit transcription of these genes (data not shown), which suggests that the genes are transcribed by RNA polymerase II. The experiments shown in Figure 1A were carried out using nuclei from excised EHS. Previously, we have shown that SAUR expression is organ-specific, with greatest expression in elongating regions of etiolated stems (McClure and Guilfoyle, 1987; B. McClure and T. Guilfoyle, manuscript submitted).

We also tested the transcriptional inducibility of these genes in primary leaves, where SAURs are expressed weakly. Figure 1B compares auxin-inducible transcription of gene 6B with another auxin-inducible soybean gene, GH3 (Hagen et al., 1984). The results show that, although transcription of this SAUR gene is strongly auxin-inducible in the EHS, only a low level of inducibility is found in primary leaves. In contrast, the GH3 gene is transcriptionally induced by auxin in both hypocotyl and leaf tissues (Hagen et al., 1984). We have obtained similar organ-specific transcription results using gene-specific cDNAs for genes 10A5 and 15A (B. McClure, unpublished results).

To investigate the kinetics of transcriptional induction of the SAUR genes, we isolated nuclei from EHS incubated in the presence of 50 μM 2,4-D for various lengths of time. The results of this experiment are shown in Figure 1C. Technical difficulties associated with studying the transcriptional response in the EHS (i.e. the time required for harvesting organ sections and preparation of nuclei) com-

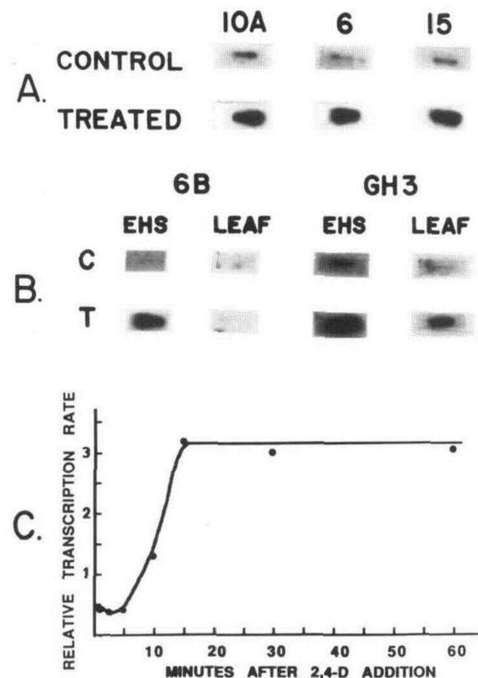

Figure 1. Transcriptional Regulation of the SAUR Genes.

(A) Each family of SAUR genes is transcriptionally regulated. Gene family-specific cDNA clones 10A15.5, 6.4, and 15.5 (10 μg/slot) were slot-blotted and hybridized (Hagen and Guilfoyle, 1985) to ^{32}P-RNA synthesized in in vitro nuclear run-on reactions (2×10^7 cpm/ml). Nuclei were isolated (Hagen and Guilfoyle, 1985) from soybean EHS preincubated for 4 hr as described (McClure and Guilfoyle, 1987), and then incubated for 1 hr in media either containing 50 μM 2,4-D (treated) or lacking 2,4-D (control). Of the cDNAs used here, clone 6.4 corresponds to gene 6B, clone 15.5 to gene 15A, and clone 10A15.5 to gene 10A5. Under the washing conditions used here, each gene-specific cDNA hybridizes only with its gene family members.

(B) Organ specificity of transcriptional regulation. The antisense strand cDNA clones 6B23 and M13-193 (5 μg/slot) were slot-blotted and hybridized (Hagen and Guilfoyle, 1985) to ^{32}P-RNA synthesized in in vitro nuclear run-on reactions (2×10^7 cpm/ml). Washing conditions were as described in "Methods." EHS nuclei were obtained as described for **(A)**. C, unincubated EHS; T, EHS preincubated 4 hr as in **(A)** and then treated with 50 μM 2,4-D for 60 min. Leaf nuclei were isolated from primary leaves preincubated overnight and then treated for an additional 60 min in KPSC media with either no hormone (C) or containing 50 μM 2,4-D (T).

(C) Transcriptional induction kinetics. Clone 6B23 (5 μg/slot) was slot-blotted and hybridized (Hagen and Guilfoyle, 1985) to ^{32}P-RNA synthesized in in vitro nuclear run-on reactions (2×10^7 cpm/ml). EHS nuclei were isolated (Hagen and Guilfoyle, 1985) from organs that were preincubated in KPSC media 4 hr (0 min) and then treated with 50 μM 2,4-D for 1 min, 2.5 min, 5 min, 10 min, 15 min, 30 min, and 60 min as indicated. Hybridization was quantitated with an Ambis scanner. Relative transcription rate reflects counts per minute hybridized to each slot blot after background subtraction.

plicate interpretation of the results for very early time points (i.e. less than 5 min). The results in Figure 1C do show, however, an auxin-induced increase in the rate of SAUR gene transcription within 10 min, using a probe for gene 6. Experiments using the other gene-specific cDNAs shown in Figure 1A show a similar sharp increase in the rate of transcription between 5 min and 10 min after auxin treatment (B. McClure, unpublished results).

Genomic Organization

Figure 2 summarizes the results of experiments designed to elucidate the organization of the auxin-regulated genes. DNA gel blot experiments using probes specific for genes 10A5, 6B, and 15A exhibit three types of hybridization patterns to soybean genomic DNA (Figure 2A). All three probes hybridize to an identical pattern of fragments from a BamHI digest, with a major band of 11.5 kb to 12 kb. Probes for genes 6B and 15A detect a common set of EcoRI fragments (i.e. the major band at 3.2 kb) that are distinct from those detected by a probe for gene 10A5 (i.e. the major band at 1.1 kb). A third type of hybridization pattern is exemplified by the HindIII digest, where the probes for genes 10A5 and 6B detect a common set of fragments (i.e. the major 9-kb HindIII fragment) and a probe for gene 15A detects a separate fragment (i.e. the 2.8-kb fragment). Hybridization of these gene-specific probes to

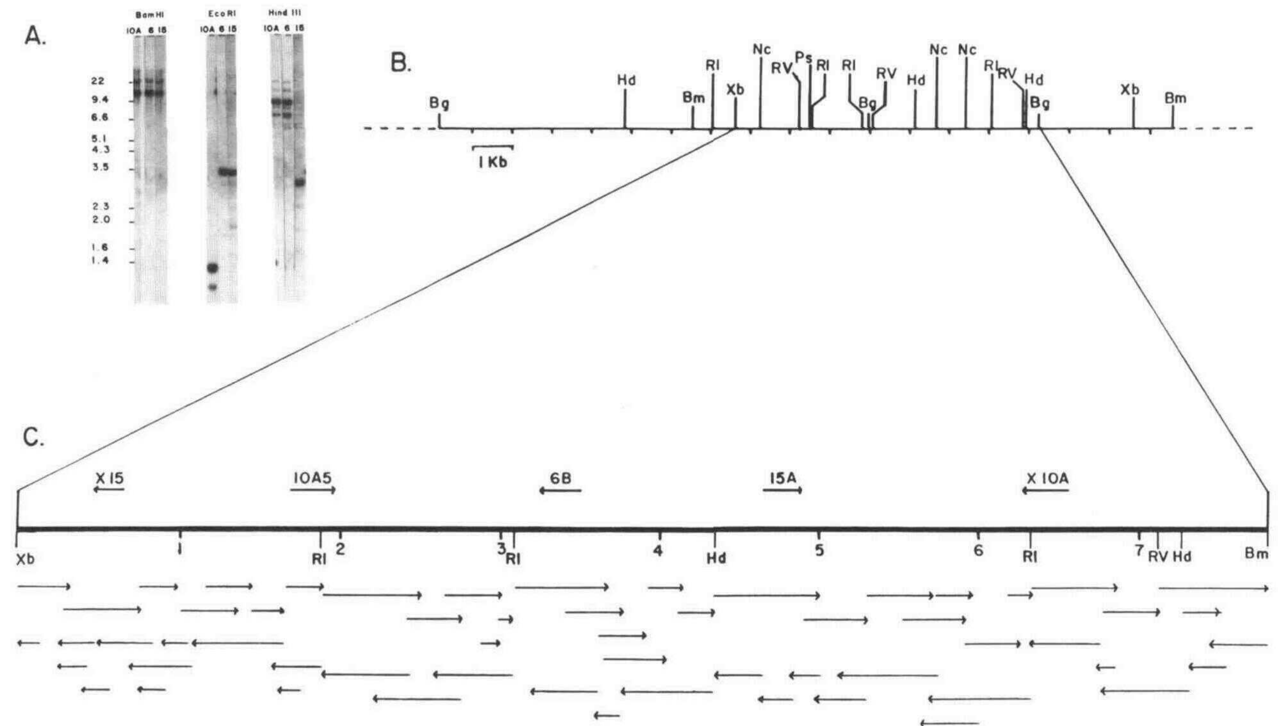

Figure 2. Organization of the SAUR Locus.

(A) Soybean genomic DNA blot analysis. Soybean genomic DNA was digested with the restriction enzymes BamHI, EcoRI, or HindIII as indicated. The digested DNAs (30 μg/lane) were separated in a 1.0% agarose gel in TAE buffer. The gel was blotted onto a nylon membrane and probed separately with the gene family-specific cDNAs 10A15.5, 6.4, or 15.5 as described in "Methods." The probes are indicated at the top of each lane. Size markers were HindIII and HindIII + EcoRI digested λ DNA. The positions and sizes (kb) of some of these marker bands are indicated at the left.

(B) Map of the SAUR locus. The positions of the restriction sites were determined by hybridization analysis of single and double restriction digests of soybean genomic DNA as described in "Methods." The enzyme sites are abbreviated as follows: BamHI, Bm; BglII, Bg; EcoRI, RI, EcoRV, RV; HindIII, Hd; NcoI, Nc; PstI, Ps; XbaI, Xb. Distances are indicated in kilobases. The precise locations of the genes 10A, 6, and 15 were determined by the presence of restriction sites which occur in both the genomic DNA and in cloned cDNAs.

(C) Sequencing strategy. The genomic subclone pXB21 was further subcloned at the BamHI, EcoRI, EcoRV, HindIII, and XbaI sites shown. Individual sequencing templates are indicated for both stands at the bottom. The positions and orientations of the five ORFs are also indicated. The figure is numbered in kilobases from the XbaI site. The ORFs correspond to base numbers 713–469 (X15), 1697–1978 (10A5), 3523–3255 (6B), 4656–4904 (15A), and 6868–6302 (X10A).

cloned fragments of soybean genomic DNA in Charon 35 show the same three types of hybridization patterns (M. Gee, unpublished results). These results are consistent with a model in which the three genes are clustered, with gene 6B occupying a central position.

The relatively simple patterns of hybridization (Figure 2A) allowed the construction of a restriction map of the gene cluster by direct hybridization analysis of soybean genomic DNA. Figure 2B summarizes the results of genomic restriction mapping experiments. The hybridization method used to construct the map shown in Figure 2B can only detect cleavage sites close to the gene cluster. The observation that a single major band occurs in many genomic restriction digests suggests that there may only be a single copy of this gene cluster per complement of soybean DNA.

DNA Sequence Analysis

The strategy for obtaining the DNA sequence of the gene cluster is shown in Figure 2C. A restriction map generated from the sequence data confirms the positions of the sites shown in Figure 2B with two exceptions. The first exception is that the genomic DNA restriction mapping results indicate that a BglII site is located at the position where the XbaI-BamHI subclone ends. The cause of this discrepancy is not known. The second exception is that the sequence data show that the NcoI site lying between genes 15A and X10A (Figure 2B) is actually a cluster of three NcoI sites. These three sites are too close together to be resolved by the DNA gel blot data.

Figure 2C shows the positions of five homologous open reading frames (ORFs), each about 240 bp to 270 bp long, that were revealed by the DNA sequence data. The genes containing these five ORFs are transcribed in alternate directions and are spaced at intervals of about 1.25 kb. The genes labeled 10A5, 6B, and 15A correspond identically with cDNA sequences (Figure 3). The genes referred to as X10A and X15 do not show sequence identity with any cDNAs yet isolated. However, the sequences of these two genes clearly place them in the 10A and 15 families. None of the SAUR genes contains introns.

An alignment of the sequences surrounding the five genes is shown in Figure 3A. This alignment has been modified from one generated with the Intelligenetics GENALIGN program to maximize the overlap of similar sequences and minimize the number of gaps in each sequence. The five genes are ordered such that sequences in adjacent lines show the greatest possible identity over the entire length of the segments shown. Thus, the two 10A family members are clustered, as are the two 15 family members. Gene 6B occupies the central position in Figure 3A because it is more closely related to both the 10A and 15 families than these two families are to each other.

The alignment in Figure 3A shows that the five genes share several regions of high homology. The locations of the homologous regions are illustrated in Figure 3B. The histogram in Figure 3B represents the numbered positions containing identical residues among the five genes within adjacent nonoverlapping five nucleotide windows based on the alignment shown in Figure 3A. Thus, a score of 5 represents complete identity at all five positions in the window, whereas a score of 0 indicates at least one mismatch at each position.

The region labeled DUE in Figure 3 is the distal upstream element. The DUE of each of the five genes contains the sequence CttgANAaAGtCCTC(Py)aAG(Pu)CA.[3] The region labeled NDE in Figure 3 is the NdeI element, which, in four of the five genes, contains the recognition sequence for the restriction endonuclease NdeI (CATATG).

The region immediately upstream of the TATA boxes and including the NDE is very rich in pyrimidine residues. Base composition analysis shows that the pyrimidine contents of the regions are as follows: 63% for gene 15A, 68% for gene X15, 68% for gene 6B, 70% for gene X10A, and 58% for gene 10A5. Three of the genes (i.e. genes 15A, X15, and 6B) contain the sequence CAAT within this pyrimidine-rich stretch. However, none of the genes contains the CAAT box consensus sequence GG(c/t)CAATCT (Schoeffl et al., 1984). The spacing between the NDE and the TATA box varies from 72 bp in gene 15A to 41 bp in gene 10A5. The TATA boxes of all of the genes except 10A5 resemble the sequence TC(TA)$_{4-5}$. These sequences are similar to the consensus reported previously (Messing et al., 1983), except that the stretch of A residues at the 3′ end of the TATA motif is missing in the SAUR genes. Gene 10A5 lacks both the TC residues before the TATA motif and the stretch of A residues at the 3′ end of the motif. Overall, the gene 10A5 sequence is the most divergent of the five genes in the TATA region and in the other upstream regions.

In addition to the 5′ sequence elements, sequence analysis of the SAUR gene cluster reveals two highly conserved elements in the 3′ regions of the genes. The first of these is referred to as the downstream element (DST, Figure 3). The DST occurs within the 3′-untranslated region of the SAUR genes and begins 11 bp to 19 bp downstream from the stop codon. The DST consists of 36 bp to 37 bp with the consensus GGAn$_{5-6}$ATAGATT aGn$_{8-9}$A(t/a)TTtGTAcA.[3] The second 3′ conserved element is referred to as the 3′-intergenic element, and the consensus sequence for the three elements shown in Figure 4 is (t/a)ATTn$_{3-4}$GGCAATA(c/t)CA(t/c)(a/g)GTA.[3] Two copies of this sequence occur as part of slightly

[3] Capitalized symbols are invariant bases and lower case symbols represent a consensus shared by all but one sequence. The symbols N, Pu, and Py represent either any nucleotide, a purine, or a pyrimidine, respectively.

Figure 3. SAUR Gene Homologies.

(A) SAUR gene sequence alignments. The most homologous regions of the five genes in the sequenced portion of the SAUR locus were aligned by the Intelligenetics program GENALIGN. The initial alignment was refined slightly to minimize the number of sequence gaps and maximize alignment of homologous segments. Identical bases in adjacent lines of sequence are capitalized and indicated by an asterisk. Sequence segments discussed in the text are boxed. The underlined regions of genes 10A5, 6B, and 15A are identical to cDNA sequences. The sequence segments shown here correspond to base numbers 985–365 (gene X15), 1428–2073 (gene 10A5), 4370–5006 (gene 6B), 4370–5006 (gene 15A), and 6868–6203 (gene X15).

(B) SAUR gene homology scores. Nonoverlapping five-character segments of the alignment shown in **(A)** were given a score of 0 to 5 according to the number of bases identical in all five SAUR genes. Distances indicated at the bottom of the figure correspond to the number of characters, including bases and gaps, in the alignment shown in **(A)**. The individual segment scores were plotted as a histogram. Scores of 4 or 5, corresponding to 1 or 0 mismatches, are indicated in black. Homologous regions discussed in the text are indicated at the top.

longer direct repeats within a region of complex symmetry located about midway between the stop codons of genes 15A and X10A (Figure 2C). The 3′-intergenic element labeled A (Figure 4) lies 741 bp downstream from the gene 15A stop codon and 657 bp downstream from the gene X10A stop codon. The adjacent copy of this element (Figure 4C) lies 772 bp downstream from the gene 15A stop codon and 636 bp downstream from the gene X10A stop codon. The other copy of the 3′-intergenic element (Figure 4B) lies 659 bp downstream of the gene 6B stop codon and 618 bp downstream of the gene 10A5 stop codon. A possible fourth copy of this element (TATatgtcaGCAATAgtcTAtcA,[4] matching the consensus at 13/19 positions) occurs 188 bp downstream from the gene 10A5 stop codon. No sequences resembling the 3′-intergenic element occur in the 463 bp of known sequence beyond the gene X15 stop codon. We do not know, however, whether a sequence homologous to this element exists further downstream of the X15 gene.

The longest stretch of high homology among the five genes is within the ORF. As shown in Figure 3A, the ORFs begin at the first occurrence of ATG in the regions of the gene sequences that are identical with cDNA sequences. The sequence context of the putative translational initiation codons (AcAAcA ATG G)[3] closely resembles that found in the highly expressed zein genes (Heidecker and Messing, 1986). Overall, about 60% of the residues are identical in the ORFs of all five genes. The most variability is seen in the 5′ third of the ORF regions where genes 15A and X15 contain 24-bp deletions, and the 10A5 gene contains a 3-bp insertion, relative to the gene 6B sequence. The ORFs in genes 15A and X15 end in a TGA codon, whereas the other three ORFs end with a TAA codon.

Translations of the ORFs found in the five genes are aligned in Figure 5. As in Figure 3A, the translations are arranged such that the most closely related sequences are adjacent. The polypeptides predicted from the SAUR gene ORFs are all broadly similar to each other, with molecular weights in the range of 9 kD to 10.5 kD. They also appear to be slightly acidic, with isoelectric points in the range of pH 5.93 to pH 6.06. Overall, the five polypeptides show identity at 56% of the amino acid residues; however, the central region of the polypeptides appears to be most highly conserved. In the region between the proline at position 27 to the proline at position 77, 75% of the amino acids are invariant among the five SAUR genes (Figure 5). Analyses of these putative polypeptides by the Chou and Fasman algorithm (Chou and Fasman, 1978), combined with hydrophobicity plot analysis (Hopp and Woods, 1982), suggest that this central region may fold into a hydrophobic three-strand β-sheet followed by a hydrophilic helical region. In spite of the high degree of overall amino acid sequence similarity among the five polypeptides, each of

[4] Capital letters indicate bases that are identical between two sequences being discussed or matching a consensus.

```
A      TATTT-GGGGCAATAcCAtgGTA
       ***** ********* **  ***
B      TATTTtGGGGCAATAtCAcAGTA
        ***    ******* ** ****
C      aATTccatGGCAATAcCAtAGTA

CON    tATTn3-4GGCAATAcCAtaGTA
       a               t  cg
```

Figure 4. Comparison of 3′-Intergenic Region Elements.

The sequence elements were aligned with the Intelligenetics program GENALIGN. The sequences labeled **A** and **C** occur as parts of 30-bp direct repeats falling approximately in the center of the 3′-intergenic region between genes 15A and X10A. The element labeled **B** occurs in the 3′-intergenic region between genes 10A5 and 6B (see Figure 2). In the alignment, identical bases in adjacent lines are indicated by capital letters and an asterisk. A consensus sequence is indicated. Capital letters denote invariant bases among the three sequences, and lower case letters denote a two-thirds consensus. The bases shown here correspond to base numbers 5645–5666 **(A)**, 2592–2570 **(B)**, and 5676–5698 **(C)**.

the three SAUR gene families shows a unique combination of distinctive features. The most striking of these features is the eight amino acid deletion characteristic of the 15 family polypeptides (i.e. 15A and X15). The 10A family polypeptides (i.e. 10A5 and X10A) are characterized by a two-amino acid extension at the carboxyl terminus. The six family sequences lack either of these two features. It is also worth noting that, although the polypeptides appear to be only slightly acidic overall, there is a distinct asymmetry to the distribution of charged amino acid side chains. Of the first 50 residues, two are acidic and nine are basic (including the amino-terminal methionine). In contrast, the carboxyl-terminal portion of the sequence shows eight acidic residues (amino acid residues 51 through the carboxyl terminus) and no basic residues.

DISCUSSION

Organization and Expression of the SAUR Locus

We have isolated a group of auxin-regulated genes that are clustered in the soybean genome. Since three of the genes in the SAUR locus have complete sequence identity with cloned cDNAs, it is clear that at least these three genes are expressed. Previously, we have shown that the RNAs transcribed from these genes accumulate very rapidly in response to auxin (i.e. within 2.5 min after auxin application [McClure and Guilfoyle, 1987]). The results in Figure 1 show that at least part of this RNA accumulation is due to increased transcription from the genes we have isolated and sequenced. It is interesting to note that the

```
              10        20        30        40        50        60        70        80        90
               !         !         !         !         !         !         !         !         !
X15   MGFRLPGI-RKAS--------nAvDAPKGYLAVYVGEKmKRFVIPVSYmNQPSFQDLLtQAEEEFGYDHPMGGLTIPCSEeVFQrITcCLN
      ******** ****         * ************** ********* ********* ********************* *** ** ***
15A   MGFRLPGI-RKAS--------KAaDAPKGYLAVYVGEKlKRFVIPVSYLNQPSFQDLLSQAEEEFGYDHPMGGLTIPCSEDVFQcITSCLN
      ******** ****        ** *  ***********  ********** **************** ** ************* *** **
6B    MGFRLPGI-RKASfsANQASSKAVDVeKGYLAVYVGEKMRRFVIPVSYLNkPSFQDLLSQAEEEFGYhHPnGGLTIPCSEDVFQhITSfLN
      ******** ** *  ******* * * **** **** ***** ******* ******* ******** ** ******* ** *   ** **
X10A  MGFRLPGI-RKTSiaANQASSKsVeVPKGYLvVYVGDKMRRFlIPVSYLNqPSFQDLLnQAEEEFGYDHPMGGLTIPCKEDEFLtVTSHLNdL
      ****  ** * **    ** ** * *****  ********** ******* **** ** ********************* *** ** *** *
10A5  MGFRiaGIvRrTSfyttQAaSKrVdVPKGYaaVYVGDKMRRFtIPVSYLNePSFQeLLsQAEEEFGYDHPMGGLTIPCKEeEFLnVTaHLNeL

CON   MGFRlpGI-RkaSf-anqasskavdvpKGYlaVYVGeKmrRFvIPVSYlNqPSFQdLLsQAEEEFGYdHPmGGLTIPCsEdvFq-iTscLN.l
```

Figure 5. Alignment of Putative SAUR Polypeptide Sequences.

The translated ORFs from genes X15, 10A5, 6B, 15A, and X10A were aligned with the Intelligenetics program GENALIGN. Identical amino acids in adjacent lines are capitalized and indicated by an asterisk. A consensus amino acid sequence is shown below the alignment. In the consensus sequence, capitalized residues denote invariant amino acid residues.

kinetics of transcriptional induction of the genes appears to lag somewhat behind the RNA accumulation response (Figure 1C; McClure and Guilfoyle, 1987). Although this may be due to the technical difficulties associated with transcription studies in the EHS, it is possible that there is regulation of SAUR expression at the posttranscriptional level. We have reinvestigated the effect of cycloheximide on SAUR expression, and, in contrast to our earlier report (McClure and Guilfoyle, 1987), we have found that, at least under certain conditions, cycloheximide is capable of inducing SAUR accumulation. Cycloheximide-induced accumulation of other auxin-responsive mRNAs has been reported (Theologis et al., 1985; Theologis, 1986). Additional results with the SAUR genes suggest that the cycloheximide effect is due to RNA stabilization rather than transcriptional induction (B. McClure, unpublished results). In other experiments, we have shown that SAUR transcripts turn over very rapidly in plants (McClure and Guilfoyle, 1989; B. McClure and T. Guilfoyle, manuscript submitted). Thus, it seems possible that, in addition to transcriptional control, the stability of auxin-inducible RNAs may be subject to auxin control. In this regard, it is interesting to note that the 3′-untranslated regions of all SAUR sequences obtained to date contain a conserved sequence element (i.e. the DST), which may serve a similar function in mRNA stabilization as an element in the 3′ region of the human *c-fos* gene (Treisman, 1985).

The sequence data presented here show that the soybean genomic locus we have mapped contains at least five genes. We have isolated and sequenced several additional soybean SAUR cDNAs (B. McClure, unpublished results) and these data suggest that at least four additional SAUR genes exist, but are not accounted for in the sequenced portion of this locus. Currently, we are examining regions that flank the sequenced portion of the SAUR locus to determine whether additional genes may be placed on the map shown in Figure 2B (G. Hagen, unpublished data). Indeed, the organization of the genes we have already characterized suggests the possibility that additional pairs of divergently transcribed genes may lie either to the left of gene X15 or to the right of gene X10A.

Upstream Sequence Elements

Alignment of the five genes we have characterized reveals several regions that may be of potential importance for the regulation of these genes. One prominent region of homology among the five SAUR genes begins about 250 bp upstream of the ATG codons. For convenience, we divide this area into two regions which we refer to as the DUE and the NDE. However, we do not wish to imply that there is necessarily any functional significance to dividing the regions in this way.

Recently, Ainley et al. (1988) have published the sequences of two auxin-regulated soybean genes, Aux28 and Aux22. These genes encode auxin-inducible RNAs, JCW1 and JCW2, that are expressed in the EHS, as are the SAUR genes. Furthermore, the specificity of the accumulation of RNAs JCW1 and JCW2 in response to different auxins, auxin analogs, other plant hormones, inhibitors, and specific toxins is very similar to that observed for the SAURs (Walker et al., 1985; McClure and Guilfoyle, 1987). However, the patterns of expression are not identical. For example, the kinetics of the auxin-induced accumulation of RNAs JCW1 and JCW2 in EHSs is much slower than that observed for the SAURs. RNAs JCW1 and JCW2 begin to accumulate about 15 min after auxin treatment (Walker and Key, 1982), whereas the SAURs have been shown to begin accumulating after 2.5 min of auxin treatment (McClure and Guilfoyle, 1987). In

spite of these differences in kinetics, one might still expect some similarities between the putative regulatory regions of the different auxin-regulated genes. We performed detailed comparisons between the sequences of the SAUR genes (i.e. sequences 1 kb upstream and downstream of the SAUR gene ORFs including the DUE, NDE, and DST) and other published auxin-regulated genes (Ainley et al., 1988; Czarnecka et al., 1988). Two potential regulatory sequences were described that occur in both the Aux22 and the Aux28 genes. The first of these, the A box (TGATAAAAG), occurs 398 bp and 475 bp upstream of the cap sites in Aux22 and Aux28, respectively. We do not find sequences related to the A box in an analogous position in the SAUR genes. The best match between the A box sequence and the 5′ regions of the SAUR genes lies in the DUE. For example, gene 6B contains the sequence TGATAAAG, which only partially overlaps the most conserved part of the SAUR gene DUE. The other four SAUR genes do not show a close match to the A box sequence.

The other potential regulatory elements discussed by Ainley et al. (1988) are the B box sequence (GGCAGCATGCA), which occurs at 295 bp and 216 bp upstream of the cap sites in Aux28 and Aux22, respectively, and the B′ sequence (GCACCATGC), which occurs 218 bp upstream of the Aux28 gene. We have further analyzed the sequence contexts of these B box elements in Aux22 and Aux28, and our results indicate that each of their B box elements lies 17 bp to 18 bp upstream of a sequence that resembles the CCCAT motif observed in the SAUR gene NDE. Although the SAUR genes do not contain B box elements per se, there is a partial overlap of the B box sequence with the NDE. The best match is between the Aux28 B′ box region (gCACCATgcGtCCt-TGTaTtataCGtcCCCtT)[4] and the gene 6B NDE region (aCACCATatGcCCcTGTcTctgtCGgtCCCaT).[4] The other four SAUR genes show less homology with the B box regions, but in each case the best match is with the variant B′ box element.

Like the Aux28 gene, which contains a partial repeat of one of the conserved sequences (i.e. the B and B′ box elements [Ainley et al., 1988]), the 5′-intergenic region between genes 6B and 15A also contains a partially repeated upstream element. A portion of the NDE is repeated 398 bp upstream of, and in the same orientation as, the NDE in gene 6B. It consists of the sequence CCATATGCCCTTGTtTt,[4] but the CCCAT motif is not present. Due to the spacing of genes 6B and 15A, this partial repeat of the NDE lies in a similar position relative to these two genes (i.e. 304 bp upstream of the gene 15A NDE).

We have also compared the upstream sequences of the SAUR genes with the gene 54 sequence published by Czarnecka et al. (1988). Gene 54 is auxin-inducible (Czarnecka et al., 1984; Hagen et al., 1984) and regulated at the level of transcription (Hagen and Guilfoyle, 1985; Hagen et al., 1988). We do not find any significant sequence homology between gene 54 and the SAUR genes. Perhaps this is due to the fact that, in addition to being regulated by auxin, gene 54 is regulated by a variety of other factors including heat shock and heavy metal stresses (Czarnecka et al., 1984; Hagen et al., 1988). Such complex regulation may involve a common element responsive to many stresses or may require interaction of many overlapping, upstream sequence elements. These factors may prevent detection of homologous putative auxin-regulatory sequences by computer analysis alone.

In addition to detailed comparisons between the SAUR gene sequences and the published auxin-regulated sequences, we have also searched the plant and fungal sequences in the National Institutes of Health (NIH) Genbank (1.5×10^6 bp, release 55) for sequences homologous to the DUE and NDE. This search did not reveal any likely sequence counterparts to the SAUR gene upstream elements.

Downstream Sequence Elements

The sequences of the five SAUR genes show a high degree of homology in the 3′-untranslated regions. The sequence homology in the 3′-untranslated region has been detected in all 10 soybean SAUR cDNAs sequenced to date, as well as a cDNA from pea (B. McClure, unpublished). The DST appears to be unique to the SAURs, since an homologous sequence does not occur in any of the published auxin-regulated genes (Ainley et al., 1988; Czarnecka et al., 1988), nor are similar sequences found in the plant and yeast sequences of the NIH Genbank data base.

A region of potential importance for the expression of the SAUR genes is the 3′-intergenic element ((t/a)ATTn$_{3-4}$GGCAATA(c/t)CA(t/c)(a/g)GTA,[3] Figure 4), which is located about 650 bp to 750 bp downstream from the ORFs of genes 10A5, 6B, 15A, and X10A. There are no regions of significant homology to this element in the 3′ regions of other published auxin-regulated genes. A search of the plant and yeast sequences in the NIH Genbank data base revealed one sequence with fairly high homology to this element. The soybean heat shock gene 6834 contains the sequence AtgTtagGGCAATACaATGaTA[4] located about 150 bp downstream of the stop codon (Schoeffl et al., 1984). Although this sequence is very similar to the SAUR 3′-intergenic element consensus, its distance from the stop codon is quite different. The spacing of this heat shock sequence from the stop codon more closely resembles that of the degenerate 3′-intergenic element that occurs 188 bp downstream of the gene 10A5 ORF.

Open Reading Frames

The translations of the SAUR ORFs predict that the genes encode proteins of 9 kD to 10.5 kD. The amino acid

sequences of these putative proteins are not highly homologous with any other published amino acid sequences. The absence of any apparent signal sequence in the putative proteins suggests that they would be localized within the cytosol rather than extracellularly or compartmentalized in an organelle such as the chloroplast or mitochondria. There are no obvious features of these sequences that suggest a function for the SAUR proteins. Comparison of the five amino acid sequences reported here with three other soybean SAUR sequences and one pea sequence (B. McClure, unpublished data) confirms that the N-terminal third and the extreme C terminus of the ORFs are the most subject to amino acid substitutions (B. McClure, unpublished data). Thus, it seems likely that the central conserved portion of these proteins is most important for whatever function they fulfill.

In conclusion, we have characterized a family of auxin-regulated genes that are tightly clustered in the soybean genome. We have evidence that at least three of the genes we have sequenced are expressed and are regulated at the level of transcription. The regulatory functions, if any, of the upstream and downstream sequence elements described remain to be determined. However, given the similarities in the patterns of expression of these genes, it seems likely that at least some of these conserved sequences are important for the regulation of the SAUR genes. The lack of clear homology between the conserved sequence elements in the SAUR genes and other auxin-regulated genes suggests that different factors may be responsible for the induction of different auxin-regulated genes. In view of the variety of auxin effects on plant growth and development, it seems likely that there are several mechanisms for auxin-regulated transcription. Currently, we are exploring the possible relevance of these sequences to the expression of the SAUR genes using deletion analysis, electroporation of natural and in vitro mutated DNA into protoplasts, and insertion of the SAUR genes into petunia plants (C. Brown, unpublished data). Recent results strongly suggest that the portion of the SAUR locus we have sequenced contains all the *cis*-acting sequences necessary for auxin inducibility of these genes when transferred to petunia (R. Wright and T. Guilfoyle, manuscript in preparation). Therefore, further analysis of this locus should elucidate features of these genes that are required for auxin-regulated expression.

METHODS

Plant Material

Soybean (*Glycine max* cv Wayne) seedlings were grown in darkness at 25°C in a 1:1 mixture of perlite:vermiculite as described (McClure and Guilfoyle, 1987). For the isolation of nuclei, seedlings were grown for approximately 72 hr, and the elongating section of the hypocotyl was excised. Excised EHSs were preincubated in KPSC media (10 mM KH_2PO_4, pH 6, 2% sucrose [w/v], 50 μg/ml chloramphenicol) for 4 hr, with media changes after 1 hr and 2 hr of preincubation. Auxin treatments were initiated by the addition of 10 mM 2,4-dichlorophenoxyacetic acid (2,4-D) to a final concentration of 50 μM. Soybean embryonic axes were obtained from a waste product of commercial soybean processing (kindly provided by Edible Soy Products, Hudson, IA) by retention on a number 16 standard sieve and subsequent flotation on cyclohexane/carbon tetrachloride (Guilfoyle et al., 1986).

Nuclear Run-On Transcription

Nuclei were isolated from EHSs of 72 hr etiolated soybean seedlings and used for in vitro nuclear run-on transcription assays as described by Hagen and Guilfoyle (1985). Reactions were performed using 50 μg of nuclear DNA and 130 μCi to 160 μCi of α-^{32}P-UTP (800 Ci/mmol, Du Pont-New England Nuclear, Boston, MA) per reaction. In the experiment shown in Figure 1A, strand-specific cDNAs were slot-blotted onto a nylon membrane (Zetaprobe, Bio-Rad, Richmond, CA) and hybridized as described (Hagen and Guilfoyle, 1985). In the experiment shown in Figure 1, B and C, the antisense strand of cDNA clone 6B23 was similarly slot-blotted and hybridizations were carried out as described below. For quantitation, slot blots were analyzed with an Ambis scanner (Ambis Systems Inc., San Diego, CA).

Genomic Restriction Mapping

Nuclei were isolated from imbibed soybean embryonic axes as described (Hagen and Guilfoyle, 1985), and digested with 100 μg/ml proteinase K (Boehringer Mannheim, Indianapolis, IN) in 0.5% SDS (w/v), 10 mM Tris, pH 8, 1 mM EDTA at 65°C for several hours until clear. Genomic DNA was obtained by extraction of the digest with phenol and phenol:chloroform:isoamyl alcohol (25:24:1). The DNA was dialyzed extensively against 10 mM Tris, pH 8, 1 mM EDTA. Genomic restriction fragments were separated in 0.8% or 1.0% agarose gels run with TAE buffer (Maniatis et al., 1982). Gels were run at 0.6 V/cm until the bromphenol blue tracking dye neared the end of the gel. DNA gel blots (Southern, 1975) were prepared using the alkaline blot technique of Reed (Reed and Mann, 1985) on nylon membranes (Zetaprobe, Bio-Rad). DNA gel blots were hybridized with cDNA probes labeled by the primer extension technique (Hu and Messing, 1982). Hybridization was in 1.5 × SSPE (0.27 M NaCl, 0.015 M NaH_2PO_4, pH 7.4, 1.5 mM EDTA), 1% powdered milk (w/v), 1% SDS (w/v), 0.5 mg/ml denatured salmon sperm DNA at 68°C. Stringent washes were conducted at 65°C in 0.2 × SSC (SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7), 1% SDS (w/v). A genomic restriction map was prepared by analyzing all possible single and double digests with the enzymes BamHI, BglII, EcoRI, EcoRV, HindIII, and XbaI. In addition, digests with PstI, PstI + BglII, NcoI, NcoI + BglII, NcoI + EcoRI, NcoI + EcoRV, and NcoI + HindIII were performed and analyzed. The cDNA probes used in these analyses were the 100-bp to 200-bp gene-specific cDNAs described previously (McClure and Guilfoyle, 1987) as well as nearly full-length versions of the same sequences (B. McClure, unpublished data). Since the restriction map was constructed by hybridization analysis with digests of soybean genomic DNA, only

fragments hybridizing to one or more of the cDNAs were detected. This method of analysis resulted in a map that was densely populated with sites close to the gene cluster. Sites further from the cluster, which do not result in fragments that hybridize to at least one of the probes, could not be detected by this method.

Genomic Clone Isolation

A library of soybean genomic fragments cloned into Charon 35 (Loenen and Blattner, 1983) was obtained from Ron Nagao, University of Georgia. The library was constructed by Jerry Slightom (Agrigenetics Corporation, Madison, WI) from *Glycine max* cv Wayne DNA partially digested with Sau3A to yield an average insert size of 15 kb. Approximately 5×10^5 recombinants (greater than three genome equivalents) were screened by hybridization (Benton and Davis, 1977) to cDNA inserts derived from cDNA clones 10A15.5, 6.4, and 15.5 (McClure and Guilfoyle, 1987). On the basis of restriction mapping, the λ clones isolated from the library were found to represent 11 different recombinants (M. Gee, unpublished results). One λ clone, clone 21, hybridized to all three of the cDNA inserts used to screen the library. A 7.8-kb XbaI to BamHI fragment of clone 21 containing the region hybridizing to cDNAs 10A15, 6.4, and 15 was subcloned into a derivative of pIBI25 (International Biotechnologies, Inc., New Haven, CT) to generate pXB21.

DNA Sequence Analysis

The DNA sequence of the 7791-bp insert in pXB21 was determined according to the sequencing strategy shown in Figure 2C. For sequencing, the 7791-bp BamHI-XbaI fragment was subcloned at the restriction sites shown in Figure 2C. Further sequencing templates were generated by the method of Dale et al. (1985) and synthetic primers were employed as needed. DNA sequencing reactions were run using the dideoxy chain termination method of Sanger (Sanger et al., 1977) following the Sequenase protocol (United States Biochemical Corp., Cleveland, OH). Analyses of the DNA sequence data were performed utilizing the BIONET computer resource and the Intelligenetics programs GEL, SEQ, PEP, GENALIGN, and IFIND (Kristofferson, 1987). Data base searches were made using the FASTN program of Lipman and Pearson (Lipman and Pearson, 1985) as implemented on BIONET.

Nomenclature

Based on the sequence data presented here, and on additional sequences of related cDNAs not discussed in this paper (B. McClure, unpublished data), it appears that each of the three cDNAs described previously (McClure and Guilfoyle, 1987) defines a family of closely related sequences. We refer to these as the 6, 10A, and 15 families. The RNAs in the three families are all very similar and are referred to as small auxin up RNAs (SAURs). The gene cluster described here is referred to as the SAUR locus. The three SAUR families are defined by sequence homology with the original cDNAs (McClure and Guilfoyle, 1987) and by the presence or absence of characteristic insertions and deletions in the ORFs. Thus, the names of newly discovered sequences, including the genes described in this paper, contain one of the three SAUR family designations.

ACKNOWLEDGMENTS

We thank the members of J. L. Key's laboratory for communication of their sequence data prior to publication. We thank T. Randall for his technical assistance. Computer analyses presented in this paper were performed utilizing the BIONET computer resource, which is supported by National Institutes of Health grant P41RR01685. This work was supported by grant DCB 8517676 from the National Science Foundation (to T.J.G.). C.S.B. was supported by the Food for the 21st Century Program, College of Agriculture, University of Missouri. M.A.G. was supported by a University of Missouri Molecular Biology Program graduate fellowship.

Received December 2, 1988.

REFERENCES

Ainley, W.M., Walker, J.C., Nagao, R.T., and Key, J.L. (1988). Sequence and characterization of two auxin-regulated genes from soybean. J. Biol. Chem. **263,** 10658–10666.

Baulcombe, D.C., and Key, J.L. (1980). Polyadenylated RNA sequences which are reduced in concentration following auxin treatment of soybean hypocotyl. J. Biol. Chem. **255,** 8907–8913.

Benton, W.D., and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196,** 180–182.

Chou, P.T., and Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47,** 45–148.

Czarnecka, E., Edelman, L., Schoeffl, F., and Key, J.L. (1984). Comparative analysis of physical stress responses in soybean seedlings using cloned heat shock cDNAs. Plant Mol. Biol. **3,** 45–58.

Czarnecka, E., Nagao, R.T., Key, J.L., and Gurley, W.B. (1988). Characterization of Gmhsp26-A, a stress gene encoding a divergent heat shock protein of soybean: Heavy-metal-induced inhibition of intron processing. Mol. Cell. Biol. **8,** 1113–1122.

Dale, R.M.K., McClure, B.A., and Houchins, J.P. (1985). A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: Application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13,** 31–40.

Evans, M.L. (1974). Rapid responses to plant hormones. Annu. Rev. Plant Physiol. **25,** 195–223.

Evans, M.L. (1985). The action of auxin on plant cell elongation. CRC Crit. Rev. Plant Sci. **2,** 317–365.

Evans, M.L., and Ray, P.M. (1969). Timing of the auxin response in coleoptiles and its implications regarding auxin action. J. Gen. Physiol. **53,** 1–20.

Guilfoyle, T.J. (1986). Auxin-regulated gene expression in higher plants. CRC Crit. Rev. Plant Sci. **4,** 247–276.

Guilfoyle, T.J., Suzich, J.A., and Lindberg, M. (1986). Synthesis of 5S rRNA and putative tRNAs in nuclei isolated from wheat embryos. Plant Mol. Biol. **7,** 95–104.

Hagen, G. (1987). The control of gene expression by auxin. In Plant Hormones and Their Role in Plant Growth and Development, P.J. Davies, ed (Dordrecht: Martinus Nijhoff), pp. 149–163.

Hagen, G., and Guilfoyle, T.J. (1985). Rapid induction of selective transcription by auxins. Mol. Cell. Biol. **5,** 1197–1203.

Hagen, G., Kleinschmidt, A., and Guilfoyle, T. (1984). Auxin-regulated gene expression in intact soybean hypocotyl and excised hypocotyl sections. Planta (Berl.) **162,** 147–153.

Hagen, G., Uhrhammer, N., and Guilfoyle, T.J. (1988). Regulation of expression of an auxin-induced soybean sequence by cadmium. J. Biol. Chem. **263,** 6442–6446.

Heidecker, G., and Messing, J. (1986). Structural analysis of plant genes. Annu. Rev. Plant Physiol. **37,** 439–466.

Hopp, T.P., and Woods, K.R. (1982). Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78,** 3824–3828.

Hu, N., and Messing, J. (1982). The making of strand-specific M13 probes. Gene (Amst.) **17,** 271–277.

Key, J.L. (1969). Hormones and nucleic acid metabolism. Annu. Rev. Plant Physiol. **20,** 449–473.

Kristofferson, D. (1987). The BIONET electronic network. Nature (Lond.) **325,** 555–556.

Lipman, D.J., and Pearson, W.P. (1985). Rapid and sensitive protein similarity searches. Science **227,** 1435–1441.

Loenen, W.A.M., and Blattner, F.R. (1983). Lambda Charon vectors (Ch32, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. Gene (Amst.) **26,** 205–215.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

McClure, B.A., and Guilfoyle, T.J. (1987). Characterization of a class of small auxin-inducible polyadenylated RNAs. Plant Mol. Biol. **9,** 611–623.

McClure, B.A., and Guilfoyle, T.J. (1989). Rapid redistribution of auxin-regulated RNAs during gravitropism. Science **243,** 91–93.

Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J., and Rubenstein, I. (1983). Plant gene structure. In Genetic Engineering of Plants, T. Kosuge, C.P. Meredith, and A. Hollaender, eds (New York: Plenum Publishing Corp.), pp. 211–227.

Reed, K.C., and Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13,** 7207–7221.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74,** 5463–5467.

Schoeffl, F., Raschke, E., and Nagao, R.T. (1984). The DNA sequence analysis of soybean heat-shock genes and identification of possible regulatory promoter elements. EMBO J. **3,** 2491–2497.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98,** 503–517.

Theologis, A. (1986). Rapid gene regulation by auxin. Annu. Rev. Plant Physiol. **37,** 407–438.

Theologis, A., Huynh, T.V., and Davis, R.W. (1985). Rapid induction of specific mRNAs by auxin in pea epicotyl tissue. J. Mol. Biol. **183,** 53–68.

Treisman, R. (1985). Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5′ element and c-*fos* 3′ sequences. Cell **42,** 889–902.

Vanderhoef, L.N. (1980). Auxin-regulated elongation: A summary hypothesis. In Plant Growth Substances 1979, F. Skoog, ed (Berlin: Springer-Verlag), pp. 90–96.

van der Zaal, E.J., Memelink, J., Mennes, A.M., Quint, A., and Libbenga, K.R. (1987). Auxin-induced mRNA species in tobacco cell cultures. Plant Mol. Biol. **10,** 145–157.

Walker, J.C., and Key, J.L. (1982). Isolation of cloned cDNAs to auxin-responsive poly(A)$^+$ RNAs of elongating soybean hypocotyl. Proc. Natl. Acad. Sci. USA **79,** 7185–7189.

Walker, J.C., Legocka, J., Edelman, L., and Key, J.L. (1985). An analysis of growth regulator interactions and gene expression during auxin-induced cell elongation using cloned complementary DNAs to auxin-responsive messenger RNAs. Plant Physiol. **77,** 847–850.