The sequence of the *tms* transcript 2 locus of the *A. tumefaciens* plasmid pTiA6 and characterization of the mutation in pTiA66 that is responsible for auxin attenuation

Daniela Sciaky¹ and Michael F. Thomashow²

¹Biology Department, Brookhaven National Laboratory, Upton, NY 11973, and ²Department of Bacteriology and Public Health, Washington State University, Pullman, WA 99164, USA

Received 2 November 1983; Revised and Accepted 20 December 1983

ABSTRACT

The incorporation of Ti plasmid sequences, the T-DNA, into the genomes of dicotyledonous plants causes the formation of tumors. Here we report the nucleotide sequence of one of the T-DNA "oncogenes", the transcript 2 gene of pTiA6 and we further characterize the 2.7 Kb element that has spontaneously inserted into this gene in plasmid pTiA66. The results indicate that the transcript 2 portion of the T-DNA has an open reading frame that could encode a polypeptide of 49.8 Kd. The open reading frame is surrounded by sequences that typically have roles in eucaryotic gene expression. Nucleotide sequence and Southern blot analysis also indicates that the 2.7 Kb insert in the transcript 2 gene of pTiA66 is located within the coding sequence of the gene and suggests that the element is an insertion sequence. We designate this element, IS66.

INTRODUCTION

Agrobacterium tumefaciens infects a wide range of dicotyledonous plants and causes the formation of crown gall tumors. Tumor tissue grown axenically in culture differs from normal untransformed tissue in that the transformed tissue is phytohormone independent, i.e. unlike untransformed plant cells, tumor tissue does not require the addition of auxins and cytokinins to growth media for in vitro propagation (1). It is believed that this change in phenotype is responsible for the uncontrolled proliferation of plant cells in crown gall tumors. The molecular basis of this Agrobacterium -induced change in plant cell physiology is partly understood.

All virulent A. tumefaciens harbor one of a diverse group of tumor-inducing (Ti) plasmids. During the course of infection a portion of the Ti plasmid, the T-DNA, is stably transferred to the plant cells where it becomes integrated into the nuclear genomes (2-7). Genetic and transcription analysis of the T-DNA regions of the pTiA6 and pTiC58 plasmid families have defined at least five highly conserved genes that are involved in tumor formation (8-14). Two of the "oncogenes", those encoding transcripts 1 and

2, are believed to be responsible for bringing about the auxin autotrophic phenotype. A third gene encoding transcript 4, is believed to be involved in bringing about cytokinin independence. The roles of the other oncogenes are essentially unknown.

The Ti plasmid pTiA66 is a spontaneous variant of pTiA6 (15). Whereas pTiA6 incites unorganized tumors on *Nicotiana tabacum* (i.e. the tissues are devoid of differentiated structures), pTiA66 incites tumors which in tissue culture produce shoots (16). These tumors resemble the pTiA6 *tms* mutants described by Garfinkel et. al. (8); they and others (9-11,13) have shown that mutations in the T-DNA region encoding transcripts 1 and 2 result in the formation of shooting tumors on various host plants. Indeed, when the pTiA66 T-DNA region was analyzed, it was found to have a 2.7 Kb insert that mapped to the region of the transcript 2 sequences (16).

As a step towards understanding the role(s) of transcript 2 in crown gall tumor formation and the nature of the pTiA66 2.7 Kb insert, we determined the nucleotide sequence of the transcript 2 region of pTiA6 and the sequence at the site of the pTiA66 2.7 Kb insert. The results indicate that this portion of the T-DNA region has an open reading frame that could encode a polypeptide of 49.8 Kd. The open reading frame is surrounded by sequences similar to those that typically have roles in eucaryotic and procaryotic gene expression. Analysis of the predicted amino acid sequence of the peptide suggests that the gene encodes a soluble protein.

Furthermore, the data demonstrate that the pTiA66 2.7 Kb insert is located within the transcript 2 open reading frame and that it has sequence characteristics resembling procaryotic IS elements.

MATERIALS AND METHODS

Materials

³²P-labeled α -deoxyribonucleotide triphosphates (800Ci/mmol) as well as ³²P-labeled γ -adenosine triphosphate (2000-3000 Ci/mmol) were obtained from New England Nuclear. Deoxyribonucleotide triphosphates and dideoxyribonucleotide triphosphates were obtained from P-L Biochemicals with the exception of dITP which was obtained from Sigma. Synthetic primer (15 bases) was obtained from New England BioLabs.

Polynucleotide kinase and the Klenow fragment of DNA polymerase I were obtained from either Boehringer Mannheim or New England BioLabs. IPTG was purchased from Sigma, X-gal from Vega Biochemicals and all restriction endonucleases were purchased from Bethesda Research Laboratories (BRL), New England BioLabs, or Boehringer Mannheim.

Bacterial Strains

Bacteriophage M13 strains mp8 and mp9 and the host Escherichia coli JM103 were obtained from Nina Agabian (University of Washington, Seattle, Wa.). E. coli DH1 (a gift of T.J. Kwoh) was used as a recipient for pBR322 and pBR328 constructions. A. tumefaciens strains A6 and A66 have been described previously (16).

DNA Isolation

Plasmid DNA from E. coli was isolated as described previously (16). Single-stranded phage DNA was isolated as described in Messing et. al. (17). Total DNA from A. tumefaciens strains A6 and A66 was isolated by a modification of the Marmur procedure (18). Ti plasmid DNA was isolated as described previously (16).

Cloning

Overlapping segments of DNA spanning the region of pTiA6 contained in Hind III fragment 22e and Eco RI fragment 32g (Figure 1) as well as a 435 bp Bgl II/Sal I fragment and a 460 bp Bgl II/Sma I fragment comprising the junctions between the 2.7 Kb insert and the surrounding T-DNA (19, D. Sciaky, unpublished results) were inserted into M13 mp8 and mp9 vehicles developed by Messing and Vieira (20). Other recombinant molecules used in this study were: (a) pDS236-1, containing a 560 bp Sal I fragment from the internal portion of the 2.7 Kb insert cloned in pBR322; (b) pDS180-1 containing the Eco RI 32g fragment of pTiA6 cloned in pBR328; and (c) pDS177-4 containing the Eco RI 32g fragment and the 2.7 Kb insert from pTiA66 cloned in pBR328.

DNA Sequence Analysis

The chain termination method (21) was used primarily with or without the modification of Barnes and Bevan (22). Occasionally the cloned insert was subcut with other restriction enzymes and the resulting fragments purified and used as primers on the same templates. The chemical degradation method (23) was also used on these internal primers.

Southern Blots

Southern blots were prepared and hybridization was carried out at $T_m -17^{\circ}\text{C}$ (24). Nick-translated probe was prepared as previously described (2) except 400 pmoles of dATP, dCTP, and TTP and 900 pmoles of dGTP per μg of probe were used in the reaction. Specific activities of the probes ranged from $0.5-2.0 \times 10^8 \text{ cpm}/\mu\text{g}$.

Computer Programs

The sequence was assembled and analyzed using the programs previously

Figure 1

Map of region encoding transcript 2 and the location of the pTiA66 2.7 Kb insertion sequence (IS66). Note that the pTiA66 IS is not drawn to scale. The Hind III and Eco RI fragment nomenclature is that of de Vos, et. al. (51).

described (25-28). The programs were run on either a Digital Equipment Corporation PDP 11/44 or a VAX/VMS.

RESULTS

Nucleotide sequence of pTiA6 region encoding transcript 2.

The region of pTiA6 that encodes transcript 2 and the site of the 2.7 Kb insert present in pTiA66 is shown in Figure 1. The nucleotide sequence of pTiA6 included in Hind III fragment 22e and Eco RI fragment 32g was determined (Fig. 2). Analysis of the sequence reveals a 1404 nucleotide open reading frame beginning with a codon for methionine at position 88 and ending with a TAA termination codon at position 1489. The 5' to 3' direction of this reading frame has the same orientation as the 1.6 Kb polyadenylated mRNA that is transcribed from this section of the T-DNA in plant cells (11; M. Thomashow, unpublished results). Computer assisted scans of the sequence do not reveal any other open reading frames of significant length for either DNA strand; the next largest open reading frame is 243 nucleotides in length.

Since this region of the T-DNA is transcribed in plant cells, we expected to find transcriptional signal sequences typical of eucaryotes to flank the open reading frame; this was the case. The sequence TATATTT, which appears between positions 42 and 48, is similar to the consensus Goldberg-Hogness promoter sequence TATA^{AAA} (29). A sequence that matches the consensus "CAT" box (30), CCAAT, is also present at positions 10-15. The "CAT" box, which is associated with transcription initiation (31), is usually some 50 bp upstream from the TATA sequence but in the region of transcript 2 it is only 27 bp 5' to the TATA sequence. McKnight (32), however, has shown that this sequence in yeast does have some

GAATCCCAACAATAATGGCCCAAGCTGGTTCAGCTTGGTATATTTATTTGGCTCTCAATGGGTTTGAATTTCCAACTCAGAGCATG ^{Met}
 10 20 30 40 50 60 70 80 90
 ValAlaIleThrSerLeuAlaGlnSerLeuGluHisLeuLysArgLysAspTyrSerCysLeuGluLeuValGluThrLeuIleAlaArg
 GTGGCATTACCTCGTTAGCCCAAGCCTAGAACACCTGAACGGGAAGACTACTCTCGCTTAGAAGTCTAGAACTCTGTAGCCCGT
 100 110 120 130 140 150 160 170 180
 CysGluAlaLysSerLeuAsnAlaLeuAlaThrAspTrpAspGlyLeuArgSerAlaLysIleAspArgHisGlyAsn
 TGTGAAGCTGCAAAATCATTAACCGCCTCTGGCTACAGACTGGATGGTTTCGGGGAAGCGCCAAATAATGATCGCCATGGAAAC
 190 200 210 220 230 240 250 260 270
 AlaGlyValGlyLeuCysGlyIleProLeuCysPheLysAlaAsnIleAlaThrGlyValPheProThrSerAlaAlaThrProAlaLeu
 CGCGAGTAGGTTTCGGGATTCACACTCTGTTTTAGGGGAACATCGCTACCGGCTATTTCCACAGCGCGCTACGCCGGCGCTG
 280 290 300 310 320 330 340 350 360
 IleAsnHisLeuProLysIleProSerArgValAlaGluArgLeuPheSerAlaGlyAlaLeuProGlyAlaSerGlyAsnMetHisGlu
 ATAAACCACTTGCCAAAGATACCATCCCGGTGCGCAAGAACTTTTTAGCTGGAGCACTGCCGGGTGCTCGGGAATATGCAATGAG
 370 380 390 400 410 420 430 440 450
 LeuSerPheGlyIleThrSerAsnAsnTyrAlaThrGlyAlaValArgAsnProTrpAsnProAspLeuIleProGlyClySerSerGly
 TTATCGTTTGAATTACAAGCAACAACATATGCCACGGGGCGGTGCGAAACCCGTGGAATCCAGATCTGTATCCAGGGGGCTCAAGCGGT
 460 470 480 490 500 510 520 530 540
 GlyValAlaAlaValAlaSerArgLeuMetLeuGlyGlyIleClyThrAspThrGlyAlaSerValArgLeuProAlaAlaLeuCys
 GGTGTGCTGCTCGGTAGCAAGCCGATTGATCTTAGCGCGCATAGGCACCGATACCGGTGCACTGTTCCCTACCGGCAGCCCTGTGT
 550 560 570 580 590 600 610 620 630
 GlyValValGlyPheArgProThrLeuGlyArgTyrProGlyAspArgIleIleProValSerProThrArgAspThrProGlyIleIle
 GCGGTAGTAGGATTTCCACCGACGCTTGCTAGATATCCGGGAGATCGGATAATACCGGTAGCCCTACCGGGACACTCCCGGAATCATA
 640 650 660 670 680 690 700 710 720
 AlaGlnCysValAlaAspValIleLeuAspArgIleIleSerGlyThrProGluArgIleProProValProLeuLysGlyLeuArg
 CGCAGTGGGTAGCCGATGTTGTAATCTCGACCGGATAATTTCCGGCACACCGGAGAAATACCCCGTCCGCTGAAGGGGCTAAGG
 730 740 750 760 770 780 790 800 810
 IleGlyLeuProThrTyrPheTyrAspAspLeuAlaAspValAlaLeuAlaGluThrThrIleArgLeuLeuAlaAsnLys
 ATCGCCCTCCCTACACCTACTTTTATGATGACCTTGATGCTGATCTGCGCCTAGCACCTGAACACGATTCCGCTCTAGCAACAA
 820 830 840 850 860 870 880 890 900
 GlyValThrPheValGluAlaAsnIleProHisLeuAspGluLeuAsnLysGlyAlaSerPheProValAlaLeuTyrGluPheProHis
 GCGCTAATCTTTGTTGAAGCTAACATTCGCCACCTTGACGAACTGAATAAGGGGCCAGCTTCCCGAGTTGCACCTCTATGAATTTCCAC
 910 920 930 940 950 960 970 980 990

AlaLeuLysGlnTyrLeuAspPheValLysThrValSerPheSerAspValIleLysGlyIleArgSerProAspValAlaAsnIle
 GCTCTAAACAGATATCTCGACGACTTTGTAAAACTGTTTCTTTCTGACGTCATCAAGGAATTCGTAGCCCTGATGACCAACATT
 1000 1010 1020 1030 1040 1050 1060 1070 1080
 AlaAsnAlaGlnIleAspGlyHisGlnIleSerLysAlaGluTyrGluLeuAlaArgHisSerPheArgProArgLeuGlnAlaThrTyr
 GCCAATGGCGAAATTGATGGACATCAAAATTTCCCAAGCTGAATAGTAAGTGGCCGCCACCTCTCTACAGCAAGACTTCAAGCCACCTAT
 1090 1100 1110 1120 1130 1140 1150 1160 1170
 ArgAsnTyrPheLysLeuAsnArgLeuAspAlaIleLeuPheProThrAlaProLeuValAlaArgProIleGlyGlnAspSerSerVal
 CGCAACTACTTCAAAGCTGAATAGATTAGATCTCTCTCCCAACAGCCCTTGGTGGCCAGACCCCATAGGTGAGGATTCCTCAGTT
 1180 1190 1200 1210 1220 1230 1240 1250 1260
 IleHisAsnGlyThrMetLeuAspThrPheLysIleTyrValArgAsnValAspProSerSerAsnAlaGlyLeuProGlyLeuSerIle
 ATCCACATGGCAGCATGCTGGACATTCAGATCTAGTGGAAATGTGGCAACCAAGCAGCAGCAGGAGGCTACTGCTGGCTTGAGCATT
 1270 1280 1290 1300 1310 1320 1330 1340 1350
 ProValCysLeuThrProAspArgLeuProValGlyMetCysIleAspGlyLeuAlaAspSerAspGlnArgLeuLeuAlaIleGlyGly
 CTGTTGGCTGACACCTGATCGCTTGCTGTGGATGGAGATCGATGGATTAGCGGATTCAGCAACGCTGTGTAGCAATCGGGGG
 1360 1370 1380 1390 1400 1410 1420 1430 1440
 AlaLeuGluGluAlaIleGlyPheArgTyrPheAlaGlyLeuProAsn***
 GCATTGGAAGAGCCATTGCGATTTTGGCGGTTACCCAAATTAACCTTTCTACCATGTTCTGTTTTTACAAATTTTCAGATTGA
 1450 1460 1470 1480 1490 1500 1510 1520 1530
 TGACAATCAATCCTTGTATTGCGTCTATGAACACAGTGCCTTATGTTATAATCGAATAAATACTTCGGATGGAGATTTTGAACAACT
 1540 1550 1560 1570 1580 1590 1600 1610 1620
 AAUAAA for transcript 2
 *
 TTAATTTATGATTTAACCAATAAAGTCTTTTGCATAATCATGTTCGATAAATAATTTTATATCATGATAAACTTTAGATATTTCTTG
 1630 1640 1650 1660 1670 1680 1690 1700 1710
 GAATAGGCATCTTCATATAACAATAATCTTTTCGAATTTAAATATATATCTTAAGCAAAATACATTTTCTCTAAATATAGGGAAT
 1720 1730 1740 1750 1760 1770 1780 1790 1800
 ATAATTAAGTTCAAGCAATCTTATTAGGCATAAAGTAGTTAATACAAATTTATATAATATATTTGCGTCCTCCTTTACAAA
 1810 1820 1830 1840 1850 1860 1870 1880 1890
 TTTGTCAAATCTGCATTATGAATTAATGACTTTTTTGAATTTCTCGCCGTGTGAGGACAGACGTGAAGGCCACCCACCTTTTAATTTGG
 1900 1910 1920 1930 1940 1950 1960 1970 1980
 ATGCACCTCTCACCTGCTACAACATCATTTGTGAGCCATTGACGTTACGCTGTGCGGGCCCACTTGGCACACAGATAAACCCAGACT
 1990 2000 2010 2020 2030 2040 2050 2060 2070
 TACAAAAGTCAGCAATTGCAGCCAGCGCAAAATCCGCCAAGCGGTGACGAGCATTTTATGCAATTTTCTTCAGATGAGACTAGTTCAAGCTT
 2080 2090 2100 2110 2120 2130 2140 2150 2161

flexibility as far as relative spacing to the TATA sequence. Finally, at position 1640 to 1646, a point that is 150 bp downstream from the predicted termination codon, is the sequence AATAAA which matches the consensus polyadenylation sequence (33).

Whether or not the transcript 2 region is transcribed and translated in *A. tumefaciens* or has a role in the bacteria is unknown. However, the recent results of Schroder et.al.(34) suggest that this is a possibility. They have shown that the Hind III fragment 22e produces at low levels, a 49 Kd protein in *E. coli* minicells, a protein that is consistent with the open reading frame we have detected (see below). The promoter for transcription was apparently within Hind III fragment 22e. Schroder, et. al.(34) have also shown that coupled *in vitro* transcription/translation systems prepared from both *E. coli* and *A. tumefaciens* express the 49 Kd protein, albeit poorly. It was therefore of interest to determine whether the gene had sequences typical of other procaryotic transcription/translation signals.

In *E. coli* the consensus promoter sequence is composed of a -10 sequence, TATAAT, and a -35 sequence, TTGACA (the positions are relative to transcription start)(35-36). The transcript 2 gene has a "TATA" like sequence as pointed out above; it does not, however, have a -35 sequence. Another important procaryotic gene expression signal is the ribosome binding Shine-Dalgarno sequence, AGGAGG (37), which is usually some 4 to 7 bp upstream from the start codon. Inspection of the sequence shows that there is a GAG that begins 5 bp in front of the ATG and this GAG could potentially serve this function.

Predicted protein sequence.

The protein sequence predicted from the nucleotide sequence indicates that a polypeptide of 49.8 Kd could be synthesized. The average hydrophobicity of the predicted protein is 2.6 KJ per mole residue (as calculated by Gibson et.al.,ref.38), a value typical of soluble proteins. There are, however, transmembrane proteins such as the *Salmonella* aspartate and histidine transport receptors that have an average hydrophobicity

Figure 2

Nucleotide sequence of the pTiA6 region encoding transcript 2 and location of the IS66. Included is the amino acid sequence of the open reading frame found between nucleotide 88 and 1491. The "CAT" box (CCAAT at position 10 in the sequence), the "TATA" box (TATATTT at position 42 in the sequence) are underlined and the polyadenylation site, AATAAA, at position 1640 is noted with an asterisk. The arrow at position 524 indicates the site of insertion of the IS66.

Figure 3

Properties of the 49.8 Kd protein. Hydropathy was determined using the SOAP program of Kyte and Doolittle (40). The plot represents the average hydropathic index of a seven residue scan at each point in the peptide. When a region of 20 amino acids in length has an average hydropathic index of 1.6 (indicated by the dotted line), the probability is high that the sequences span the membrane (40).

characteristic of soluble proteins, but in addition, have specific regions that are very hydrophobic and long enough to span membranes (38,39). To determine whether the predicted protein had any sections that might span membranes, we examined the amino acid sequence by the method of Kyte and Doolittle (40). The scan (Fig. 3) is qualitatively similar to other soluble proteins (40). In addition the data indicate that there are no strongly hydrophobic regions (a region with an average hydropathy index of 1.6 per residue) that are 20 residues in length, the distance required to span a membrane (40). Thus we conclude that the predicted transcript 2 protein is probably not an integral or transmembrane protein, but rather is a soluble protein. We also hoped to gain insight into the biochemistry and function of this protein by screening the sequence database that has been compiled by GENBANK. However, significant homology with other sequenced genes was not detected.

Sequence at site of A66 insertion.

The nucleotide sequence at the site of the pTiA66 2.7 Kb insert indicates that the element is within the predicted protein coding sequence of the transcript (Fig.2) at nucleotide 524. In addition, 35 nucleotides

Figure 4

Sequence of the junctions between the T-DNA and the 2.7 Kb insertion of pTiA66. Insertion of IS66 has resulted in duplication of 8 bases of the T-DNA as a direct repeat. The insertion then forms an imperfect 20/21 bp inverted repeat. The arrow in the loop points to the beginning of the termination codon (TGA) found in frame with the open reading frame for transcript 2.

into the insertion element is an in frame termination codon, TGA, (Fig.4) which would presumably cause termination of the transcript 2/A66 insert hybrid peptide.

The data also indicate that an 8 bp duplication of the target sequence occurred at the site of insertion and that the ends of the 2.7 Kb insert comprise an imperfect 20/21 bp inverted repeat (Fig.4). These attributes are characteristic of procaryotic IS elements (41) and thus suggest that the 2.7 Kb insert is an Agrobacterium insertion sequence (IS66). If true, then it might be present at other positions in the Agrobacterium genome. Indeed, when a fragment internal to IS66 (pDS236-1) was used as a probe on Southern blots containing restricted total and/or Ti plasmid DNA from A. tumefaciens strains A6 and A66, bands appeared that did not comigrate with either the Bam HI number 8 fragment or the Eco RI 32g fragment containing IS66 in pTi A66 (Figs. 5 and 6). Part of this hybridization is due to homology of IS66 with the Ti plasmid (Fig. 5, ref. 19) and can be identified to include the IS element itself (arrow G, Fig.5), the fragment that IS66 element eventually inserted into, Eco RI 32g

Figure 5

Southern blot showing homology between IS66 and the Ti plasmids of *A. tumefaciens* strain A6 and A66. Clone pDS236-1 (containing pBR322 and the Sal I 560 fragment of IS66) was hybridized to a Southern blot containing Ti plasmid cleaved with Eco RI (lanes 1 and 2) and Bam HI (lanes 3-6). Lanes 1 and 4 contain pTiA6 (14ng), lanes 2 and 5 contain pTiA66 (11ng), lane 3 contains pDS180-1 (10ng; containing pTiA6 fragment Eco RI 32g in pBR328) and lane 6 contains pDS177-4 (14ng; containing pTiA66 fragment 32g+IS66 in pBR328). The arrows point to those fragments that the Sal I 560 fragment hybridizes to. These fragments include Bam HI 2 (A), 8 (C), 8+IS66 (B), 11 (D) and Eco RI 4 (E), 32g (G) and 32g+IS66 (F). The heavy band of hybridization above arrow F is due to hybridization of pBR322 to the pBR328 vector. These are the only bands of hybridization observed when pBR322 is used as a probe against a similar blot. The weak band of hybridization below arrow F is due to hybridization of the Sal I 560 fragment to an unidentifiable pTiA6 and pTiA66 Eco RI fragment. This fragment may overlap Bam HI fragment 2.

(arrow A, Fig.5), and Bam HI fragment number 11 (arrow D, Fig.5). Weaker homology with Bam HI fragment 2 (arrow A, Fig.5), sequences which overlap the T-DNA, was also detected. Waldron and Hepburn (19) have shown that many of the restriction sites within IS66 are conserved within the region of Bam HI fragment number 11 where the homology occurs. To determine whether Bam HI fragment 11 does contain a structure similar to IS66 the "ends" of the region of homology will be sequenced.

In addition to Ti plasmid homology IS66 has homology with sequences in

Figure 6

Southern blot showing homology between IS66 and Eco RI total and Ti plasmid DNA of *A. tumefaciens* strains A6 and A66. Clone pDS236-1 (contains pBR322 and the Sal I 560 fragment of IS66) was hybridized to a Southern blot containing Eco RI digested DNA. Lanes are as follows: (1) pDS180-1 (1.65 ng; containing pTiA6 fragment Eco RI 32g), (2) pTiA6 (0.3 μ g), (3) total DNA from A6 (5 μ g), (4) total DNA from A66 (5 μ g), (5) pTiA66 (0.3 μ g), (6) pDS177-4 (6 ng; containing pTiA6 fragment 32g+IS66). The arrows point to those bands in lanes 3 and 4 that correspond to homology of IS66 with the chromosomal and/or megaplasmid DNA (42) of the bacteria. The x in lane 3 denotes a band that hybridizes with pBR322. The fragment Eco RI 32g in lanes 1,2, and 3 was poorly retained on the nitrocellulose filter. Better evidence of hybridization appears in Figure 5.

the chromosomal and/or megaplasmid DNA (42) of strains A6 and A66 (Fig. 6, ref 19). Two extra bands of homology beyond that of the Ti plasmid appear in the Eco RI restriction pattern of total A6 DNA, while four bands appear in A66. There are therefore two to four copies of IS66 in the genomic and/or megaplasmid DNA of these *A. tumefaciens* strains indicating that the element has, as IS sequences are known to do (41), inserted into many places in the genome. Experiments are in progress to determine whether the homology in the chromosomal DNA is as extensive as that reported for the Bam HI fragment 11 (19).

DISCUSSION

One of the central unanswered question about crown gall disease concerns the mode(s) of action of the T-DNA oncogenes. It is clear that they bring about phytohormone independence of plant cells, but the mechanism by which this is accomplished is unknown. They could act directly by synthesizing auxin and cytokinin or their action might be indirect, e.g. perhaps they modulate the expression of plant genes involved in hormone metabolism. Answering this question will require a physical and biochemical description of the oncogenes as well as comparison of phytohormone metabolism in untransformed cells with that of crown gall cells incited by Ti plasmids with wild type and mutant oncogenes.

In this report, we present the nucleotide sequence of the pTiA6 transcript 2 oncogene and the sequence at the site of a naturally occurring mutation of this gene -the site of the 2.7 Kb insert in pTiA66. These data show that the transcript 2 region of the T-DNA has an open reading frame of 1404 nucleotides. The predicted amino acid sequence of this reading frame suggests that the gene product is a polypeptide of 49.8 Kd. Analysis of the amino acid sequence by the method of Kyte and Doolittle (40), and the average hydrophobicity of the predicted peptide suggests that the gene product is a soluble protein. The data also shows that the 2.7 Kb insert present in pTiA66 is located within the predicted structural gene encoded by transcript 2 and that the insert has sequence characteristics resembling other procaryotic IS elements; it has caused an 8 bp duplication of the target T-DNA sequences and the ends of the sequence comprise an imperfect 20/21 inverted repeat. Southern blot analysis suggests that the insert is present at additional sites in the A. tumefaciens A6 and A66 genome.

Other occurrences of spontaneous mutation by insertion of extraneous DNA in to the T-DNA have been reported. A 1.1 Kb insert (which has been designated IS60, ref.9) has inserted into the Eco RI 32g fragment just as has been observed for the A. tumefaciens insertion sequence (IS66). The IS60 mutation also appears to cause similar affects on tumor morphology as does pTiA66; i.e. shoots appear at the site of inoculation instead of a large amorphous overgrowth. Garfinkel and Nester (43) also isolated a spontaneous mutation (A1070) in the tmr locus of pTiA6NC. This mutation is due to a 1.5 Kb insertion in Sma I fragment 10c and either has little or no homology with IS66 (D. Sciaky, unpublished results).

Transcript 2 has been shown to be polyadenylated (11,14) and is

transcribed by RNA polymerase II (44). We therefore expected to find transcription signals typical of eucaryotic genes to surround the open reading frame; this was the case. A "TATA" sequence, a "CAT" box and a polyadenylation signal flank the open reading frame. Eucaryotic transcription sequences have also been found to surround the three other T-DNA genes sequenced to date, the nopaline and octopine synthase genes (45-47) and the transcript 7 gene (48). Direct evidence that these sequences actually function in directing transcription is lacking. However, if the "TATA" like sequence we detected has a role in poising RNA polymerase II for correct transcription initiation as in other eucaryotic genes, one would expect the 5' end of the transcript to begin some 25 to 30 bp downstream from the TATA sequence. H. Klee et.al. (personal communication; manuscript submitted) have recently mapped the 5' end of the RNA and it does in fact fall in the predicted area. These data suggest that transcript 2, as the other T-DNA genes sequenced to date, have a construction like other eucaryotic genes. It should be pointed out, however, that the steady state levels of the mRNAs produced from the T-DNA genes are very low. The T-DNA transcripts, combined account for less than 0.001% of the total polyadenylated message of the cell (49,50); transcript 2 is probably less than 0.0001%. Whether this low level of RNA is due to inefficient initiation or polyadenylation of transcription, or instability of the transcript (possibly due to inefficient capping of the message) remains to be determined.

The results of Schroder, et.al. (34) have suggested the possibility that the T-DNA oncogenes might be produced in A. tumefaciens and if so, might have a function in A. tumefaciens related to virulence. The nucleotide sequence indicates that there is a TATA sequence that might help promote transcription in procaryotes. In addition there are sequences upstream from the predicted start codon that might serve as a ribosome binding site. Thus the expression of the transcript 2 gene in A. tumefaciens remains an intriguing possibility.

ACKNOWLEDGEMENTS

We thank Richard J. Roberts, Benno Schoenborn, Gary Held and Lee Bulla Jr. for use of their computer facilities, Christian Burks and Walter Goad of Los Alamos National Laboratory for the GENBANK analysis, and Susan Johns for setting up and running the Kyte and Doolittle SOAP program. In addition we'd like to thank John Dunn and Willy Crockett for helping with the Maxam-

Gilbert reactions and Ron Aimes for aiding in construction of some of the clones.

The work was supported by grants to D.S. from USDA (59-2366-0-1-524-0) and DOE (DE-AC02-76CH0016) and to M.F.T. from NSF (PCM 8109351), USDA (59-2531-1-1-742) and NCI (CA 31313).

REFERENCES

1. Braun, A.C. and White, P.R. (1943) *Phytopathology* 33, 85-100.
2. Chilton, M.-D., Drummond, M.H., Merlo, D.J., Sciaky, D., Montoya, A.L., Gordon, M.P. and Nester, E.W. (1977) *Cell* 11, 263-271.
3. Chilton, M.-D., Saiki, R.K., Yadav, N., Gordon, M.P. and Quetier, F. (1980) *Proc. Natl. Acad. Sci. USA* 77, 4060-4064.
4. Thomashow, M.F., Nutter, R., Montoya, A.L., Gordon, M.P. and Nester, E.W. (1980) *Cell* 19, 729-739.
5. Thomashow, M.F., Nutter, R., Postle, K., Chilton, M.-D., Blattner, F., Powell, A., Gordon, M.P. and Nester, E.W. (1980) *Proc. Natl. Acad. Sci. USA* 77, 6448-6452.
6. Willmitzer, L., De Beuckeleer, M., Lemmers, M., Van Montagu, M. and Schell, J. (1980) *Nature* 287, 359-361.
7. Zambryski, P., Holsters, M., Kruger, K., Depicker, A., Schell, J., Van Montagu, M. and Goodman, H.M. (1980) *Science* 209, 1385-1391.
8. Garfinkel, D.J., Simpson, R.B., Ream, L.W., White, F.F., Gordon, M.P. and Nester, E.W. (1981) *Cell* 27, 143-153.
9. Ooms, G., Hooykaas, P.J.J., Molenaar, G. and Schilperoort, R.A. (1981) *Gene* 14, 33-50.
10. Leemans, J., Deblaere, R., Willmitzer, L., De Greve, H., Hernalsteens, J.P., Van Montagu, M. and Schell, J. (1982) *The EMBO Journal* 1, 147-152.
11. Willmitzer, L., Simons, G. and Schell, J. (1982) *The EMBO Journal* 1, 139-146.
12. Willmitzer, L., Dhaese, P., Schreier, P.H., Schmalenbach, W., Van Montagu, M. and Schell, J. (1983) *Cell* 32, 1045-1056.
13. Joos, H., Inze, D., Caplan, A., Sormann, M., Van Montagu, M. and Schell, J. (1983) *Cell* 32, 1057-1067.
14. Gelvin, S.B., Thomashow, M.F., McPherson, J.C., Gordon, M.P. and Nester, E.W. (1982) *Proc. Natl. Acad. Sci. USA* 79, 76-80.
15. Hendrickson, A.A., Baldwin, I.L. and Riker, A.J. (1934) *J. Bacteriol.* 28, 597-618.
16. Binns, A.N., Sciaky, D. and Wood, H.N. (1982) *Cell* 31, 605-612.
17. Heidecker, G., Messing, J. and Gronenborn, B. (1980) *Gene* 10, 69-73.
18. Marmur, J. (1961) *J. Mol. Biol.* 12, 468-487.
19. Waldron, C. and Hepburn, A.G. (1983) *Plasmid* 10, 199-203.
20. Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
22. Barnes, W.M. and Bevan, M. (1983) *Nucl. Acids Res.* 11, 349-368.
23. Maxam, A. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
24. Thomashow, M.F., Knauf, V.C. and Nester, E.W. (1981) *J. Bacteriol.* 146, 484-493.
25. Gingeras, T.R., Milazzo, J.P., Sciaky, D. and Roberts, R.J. (1979) *Nucl. Acids Res.* 7, 529-545.
26. Staden, R. (1978) *Nucl. Acids Res.* 5, 1013-1015.
27. Staden, R. (1977) *Nucl. Acids Res.* 4, 4037-4051.
28. Blumenthal, R.M., Rice, P.I. and Roberts, R.J. (1982) *Nucl. Acids Res.* 10, 91-102.

-
29. Goldberg, M. (1972) Dissertation (Stanford Univ., Stanford, Ca.)
 30. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) *Science* 209, 1406-1413.
 31. Shenk, T. (1981) in: *Curr. Top. Microbiol. Immunol.* 93, 25-46.
 32. McKnight, S.L. (1982) *Cell* 31, 355-365.
 33. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
 34. Schroder, G., Klipp, W., Hillebrand, A., Ehring, R., Koncz, C. and Schroder, J. (1983). *The EMBO Journal* 2, 403-409.
 35. Gilbert, W. (1976) in *RNA Polymerase*, eds. Losick, R. and Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.). pp 193-205.
 36. Rosenberg, M. and Court, D. (1979) *Annu. Rev. Genet.* 13, 319-353.
 37. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342-1347.
 38. Gilson, E., Higgins, C.F., Hofnung, M., Ames, G.F.-L. and Nikaido, H. (1982) *J. Biol. Chem.* 257, 9915-9918.
 39. Russo, A.F. and Koshland, Jr., D.E. (1983) *Science* 220, 1016-1020.
 40. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.
 41. Iida, S., Meyer, J. and Arber, W. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J.A. (Academic Press, N.Y.) pp 159-221.
 42. Casse, F., Boucher, C., Julliot, J.S., Michel, M. and Denarie, J. (1979) *J. General Microbiol.* 113, 229-242.
 43. Garfinkel, D.J. and Nester, E.W. (1980). *J. Bacteriol.* 144, 732-743.
 44. Willmitzer, L., Schmalenbach, W. and Schell, J. (1981) *Nucl. Acids Res.* 9, 4801-4812.
 45. Bevan, M., Barnes, W.M. and Chilton, M.-D. (1983) *Nucl. Acids Res.* 11, 369-385.
 46. De Greve, H., Dhaese, P., Seurinck, J., Lemmers, M., Van Montagu, M. and Schell, J. (1983) *J. Mol. Applied Genet.* 1, 499-511.
 47. Depicker, A., Stachel, S., Dhaese, P., Zambryski, P. and Goodman, H.M. (1982) *J. Mol. Applied Genet.* 1, 561-573.
 48. Dhaese, P., De Greve, H., Gielen, J., Seurinck, J., Van Montagu, M. and Schell, J. (1983) *The EMBO Journal* 2, 419-426.
 49. Willmitzer, L., Otten, L., Simons, G., Schmalenbach, W., Schroder, J., Schroder, G., Van Montagu, M., De Vos, G. and Schell, J. (1981) *Mol. Gen. Genet.* 182, 255-262.
 50. Schroder, G. and Schroder, J. (1982) *Mol. Gen. Genet.* 185, 51-55.
 51. De Vos, G., De Beuckeleer, M., Van Montagu, M. and Schell, J. (1981) *Plasmid* 6, 249-253.
-