Volume 11 Number 2 1983 Nucleic Acids Research

Structure and transcription of the nopaline synthase gene region of T-DNA

Michael Bevan*, Wayne M. Barnes[1§] and Mary-Dell Chilton

Department of Biology, Washington University, and [1]Department of Biological Chemistry, Washington University School of Medicine, Saint Louis, MO 63110, USA

Received 12 October 1982; Revised and Accepted 10 December 1982

ABSTRACT

We present the DNA sequence and plant-tumor transcription pattern of some 2400 base pairs from the right border region of pTi T37 DNA from the virulent Agrobacterium tumefaciens strain T37. This region includes the entire transcription unit encompassing the nopaline synthase gene, together with parts of other transcription units. The strategy used to determine the sequence also produced two opposing series of defined, asymmetric deletions across the target DNA region, some of which may serve future purposes in the exploitation of this sequence, which is known to be expressed in a wide variety of host plant tissues.

INTRODUCTION

Crown gall tumors on dicotyledonous plants are incited by strains of Agrobacterium tumefaciens that harbor tumor-inducing, or Ti, plasmids.[1,2] A small segment of the plasmid, called T-DNA, is transferred to, and stably maintained in, the tumorous plant cells.[3] T-DNA is covalently linked to plant nuclear DNA[4-9] and encodes polyadenylated transcripts[10-12] some of which have been shown to encode proteins.[13]

Axenic cultures of crown gall tumors are characterized by autonomous growth in the absence of the plant growth regulators auxin and cytokinin.[14] They also synthesize opines, novel low molecular weight derivatives of amino acids, keto acids or sugars that can serve as a sole source of carbon and/or nitrogen for the inciting Agrobacterium strain.[15,16] Opines are not synthesized by the inciting Agrobacterium strain itself.[16] The mRNA for lysopine dehydrogenase, the enzyme that synthesizes the opine octopine (N_2[D-1-carboxyethyl]-L- arginine), is encoded by T-DNA from octopine specific Ti plasmids.[17,18] There are an additional six polyadenylated transcripts found in octopine synthesizing tumors[12] that may encode polypeptides important in the synthesis of other opines, tumor morphology, and growth rate.[19,20] A very low level of nonspecific transcripts of T-DNA has been reported in Agrobacterium.[21]

0305-1048/83/1102-0369$2.00/0

The apparent functioning in higher plant cells of genes that derive from a prokaryote is surprising. One might argue either that T-DNA possesses regulatory sequences that promote transcription in eukaryotes, or that T-DNA transcripts originate in flanking plant DNA and are spliced into the observed mature transcripts.

We report here the nucleotide sequence and transcript map of the T-DNA gene that encodes nopaline synthase, the enzyme that synthesizes the opine nopaline. Our results show that the gene is preceded by T-DNA sequences that have a high degree of homology with regions of eukaryotic genes known to be transcribed by RNA polymerase II. These homologies allow us to draw the conclusion that, in the case of this gene at least, T-DNA contains sequences typical of eukaryotic promotors, and may not rely on a complex read-through mechanism from an adjacent plant promotor for expression. The nopaline synthase gene is unlike many other eukaryotic structural genes in that it lacks intervening sequences.

MATERIALS AND METHODS

Nucleic Acid Sequencing

A 3.4 kb Hind III fragment (Hind III fragment 23) that contains the nopaline synthase gene of pTi T37[22,23] was subcloned into pBR325[24,25] from a Charon 4A[26] library of pTi T37 constructed by N. Yadav. The Hind III fragment was purified from an agarose gel and ligated to M13 mWB2341[27] replicative form DNA which had been cleaved with Hind III and treated with calf intestine alkaline phosphatase. mWB2341 has sequences complementary to the 15-base universal primer followed by a unique Eco RI site, a unique Hind III site, and the *lac* operator. Recombinant phage were identified from their genome size as determined by analysis of infected bacteria by the toothpick assay.[28] Both orientations of the Hind III fragment were identified by annealling small aliquots of culture supernatants in pairs at 68° C for several hours, and electrophoresing the mixture in 0.7% agarose gels.[27] The M13 constructs carrying Hind III fragment 23 in each orientation are called mTi23.2 and mTi23.11. Replicative form of the recombinant phage DNA was prepared using the cleared lysate procedure.[29]

We developed a method that enabled us to obtain systematic sequence data from the 3.4 kb Hind III insert in mWB2341. This method is generally applicable to any insert in the appropriate strains of M13, and is described in detail in the adjacent paper.[27] Briefly, 5 ug of the recombinant M13 replicative form DNA was treated with DNase I in the presence of ethidium

bromide such that each molecule was nicked once randomly. This nick was then widened to a gap by a very limited digestion with Exonuclease III, and the exposed single-stranded DNA was digested with Bal 31 nuclease. The linear DNA molecules were then made flush-ended by treatment with the large fragment[30] of DNA polymerase I, and ligated to a hundred-fold molar excess of Eco RI linkers. After passage over a small column of Biogel A15 M to remove excess linkers, the DNA was digested with a large excess of Eco RI, phenol extracted, and ligated at a final concentration of approximately 2-5 ug/ml to promote recircularization of DNA molecules. These treatments resulted in the creation of a deletion extending from the Eco RI site in mWB2341 adjacent to the universal primer site to the point of the random DNase nick, with a (linker) Eco RI site at the deletion boundary. Transformation of these deleted circular molecules into competent E.coli allowed selection for phage that had deletions terminating in the Hind III insert DNA, since deletions extending a short distance past the insert DNA rendered the phage non-viable. The size of each deletion was easily estimated by electrophoresing a small sample of lysed, infected cells in agarose gels.[28] In this way asymmetric deletions of a range of sizes were catalogued and sequenced to give overlapping spans of data. Each span of data obtained is indicated schematically in figure 1 and exactly in figure 2.

Three sequencing experiments indicated in figure 2 did not arise from the kilo-sequencing strategy. MTi35 and mTi40 are inserts of Hind III - Sst II fragments in mWB2341 as described.[9] Clone 2-6 is the 920 bp Hpa II fragment inserted into the Acc I site of M13mp7.[31]

Sequence reactions were conducted according to Sanger et al.[32] with slight modifications[27] and the products were electrophoresed in thin 8% acrylamide/7M urea gels[33] in pH 8.8[34] Tris-borate electrophoresis buffer.

Radiolabelling and isolation of DNA fragments

For estimating the position of the 5′ end of the transcript, a 920 bp Hpa II fragment was purified from the 3.4 kb Hind III fragment 23 cloned in pBR325. Five micrograms of this isolated fragment were digested briefly with 0.5 ug of Exonuclease III at 15°C for 90 seconds to further expose the 5′ ends for efficient phosphorylation. The reaction was terminated by phenol extraction and ethanol precipitation and the fragments were dephosphorylated with 10 U of calf intestine alkaline phosphatase at 60°C for 1 hr. The reaction was terminated by phenol extraction and ethanol precipitation, and the product was rephosphorylated with T4 DNA kinase and γ-^{32}P ATP as described.[35] The kinase reaction was terminated by heating to 70°C for 10

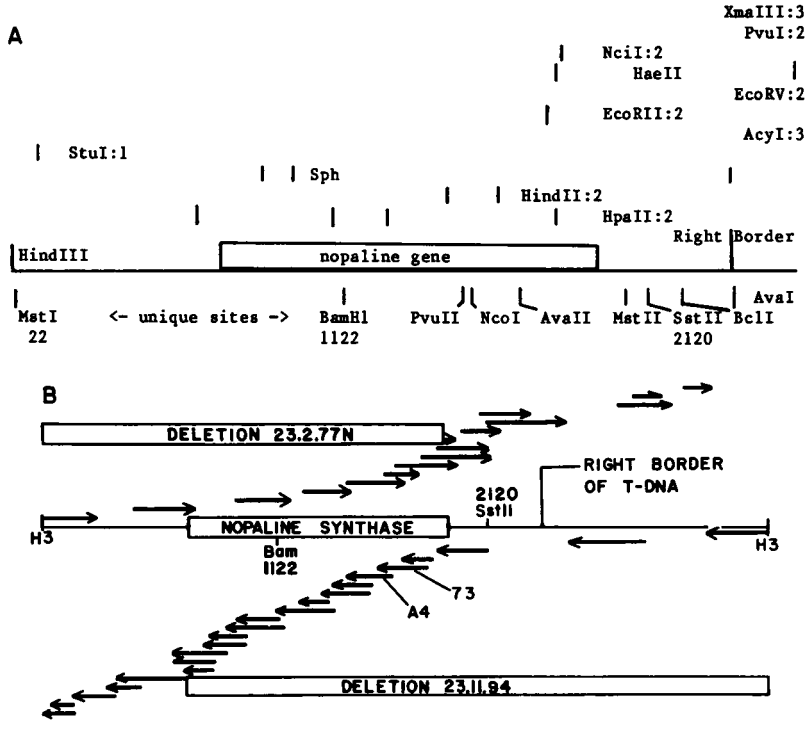


Figure 1A. Map of the target DNA region from Hind III fragment 23 of the right border region of T-DNA from pTi-T37. This map is in the standard orientation, opposite to the orientation of the sequence as presented in figure 2. The sequence is incomplete between the right border and the end of the Hind III fragment, but the number of sites encountered so far in that region are indicated after a colon (:). No sites were encountered for AccI, ApoI, BglI, BglII, EcoRI, HpaI, KpnI, MluI, NaeI, NarI, PstI, SmaI, SstI, SalI, TthII, XbaI, ShoI, XmnI.

Figure 1B. Spectrum of deletions chosen for sequencing data to cover the right border of T-DNA and the nopaline synthase gene. The arrows represent sequence data; thus the deletion endpoints are at the bases of the arrows. For scale, the positions of the unique BamHI and SstII sites are indicated, with the first base of the leftmost HindIII site as nucleotide number 1.

Three of the sequencing experiments shown are not from the kilo-deletion procedure: mTi35, mTi40, and mTi2-6 (see Methods). The experiments with mTi23.2 deletions are shown in the upper portion of the figure and are, from left to right for the deletion endpoints: mTi23.2.38N (undeleted, actually), mTi23.2.2N, mTi23.2.51N, mTi23.2.68N, mTi23.2.39N, mTi23.2.46N, mTi23.2.8, [mTi2-6], mTi23.2.64N, mTi23.2.77N, mTi23.2.68, [mTi40], mTi23.2.67, mTi23.2.49n, mTi23.2.66n, mTi23.2.67N. The mTi23.11 deletions in the lower portion of the figure are, from right to left: mTi23.11.66 (undeleted, actually), mTi23.11.65, [mTi35], mTi23.11.A1, mTi23.11.73, mTi23.11.A4, mTi23.11.A24, mTi23.11.A32, mTi23.11.A26, mTi23.11.82, mTi23.11.95, mTi23.11.86, mTi23.11.A22, mTi23.11.A8, mTi23.11.A2, mTi23.11.85, mTi23.11.A20, mTi23.11.94, mTi23.11.92, mTi23.11.96, mTi23.11.93, mTi23.11.A11. The sequence to the right, outside T-DNA, is under determined at this time.

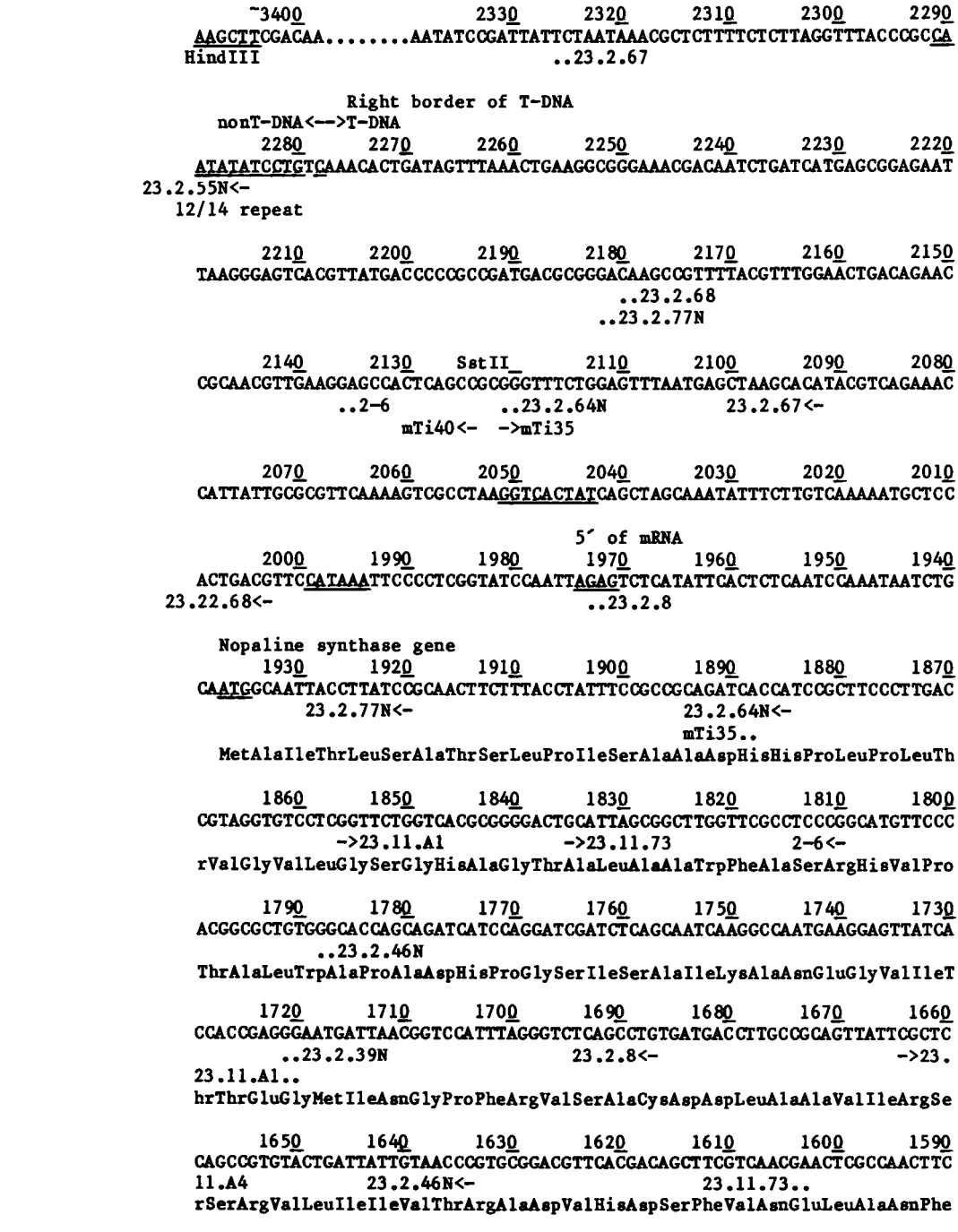
minutes, and the product was diluted to 250 ul with Sst II salts and digested with Sst II. The two labelled fragments were separated by electrophresis in a 5% acrylamide gel, and the 314 bp fragment was purified by electroelution and DEAE Sephadex column chromatography for use as a hybridization probe.

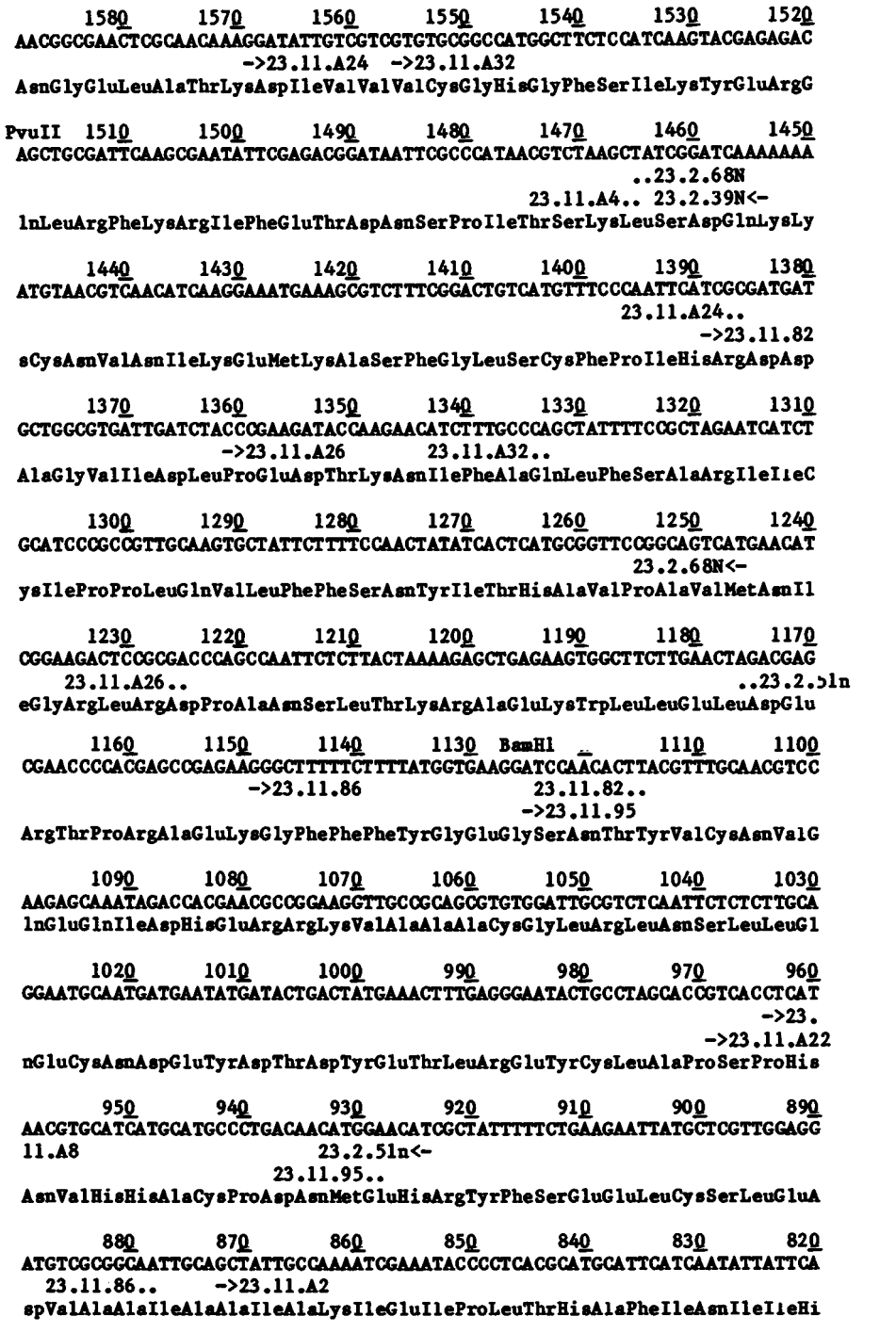
For mapping the 3′ end of the nopaline synthase transcript, the 3.4 kb Hind III fragment 23 cloned in pBR325 was cleaved with Bam HI and briefly digested with Exonuclease III as previously described. The reaction was terminated by phenol extraction and ethanol precipitation, and the exposed 5′ end was filled in with 100 uM dGTP, dTTP, dCTP and 2.5 uM α-^{32}P dATP in 50 mM NaCl, 10 mM tris HCl, pH 7.9, 10 mM $MgCl_2$, 10 mM beta-mercaptoethanol, and 2.5 units of Klenow polymerase at 37°C for 30 minutes. This enabled the 3′ ends of the DNA fragment to become radiolabelled to a high specific activity, and completely filled in as judged by electrophoresis of the radiolabelled fragment in a denaturing polyacrylamide gel. The end-filling reaction was terminated by heating the reaction to 70°C for 10 minutes, and then diluting it to 250 ul with Hind III salts. The fragments were digested with Hind III, resolved by electrophoresis in 1% agarose gels, and the 1.1 kb Hind III - Bam HI fragment complementary to the 3′ end of the nopaline synthase gene was purified by electroelution and DEAE Sephadex chromatography.

RNA isolation, hybridization and S1 nuclease digestion

RNA was isolated from freeze-dried BT37 tumor (incited by *A. tumefaciens* T37 on *Nicotiana tabacum* var. Havana 425 by Dr. A.C. Braun) tissue using the guanidine thiocyanate method.[36] Poly A+ RNA was selected by three passages over oligo-dT cellulose as described.[37] Poly A+ RNA (50 ug) was hybridized with approximately 0.1 ug of denatured end-labelled DNA fragments in 20 ul of 80% formamide, 0.4 M NaCl, 0.05 M PIPES, pH 6.4, 0.001 M EDTA AT 48°C for 8-12 hrs. Nuclease S1 reactions were initiated by rapidly diluting the hybridization mixture into 10 volumes of ice-cold S1 nuclease buffer containing an appropriate amount of S1 nuclease. The reaction was brought to room temperature (approximately 20-22°C) and incubated for 2 hrs, after which it was terminated by phenol extraction and ethanol precipitation. Exonuclease VII reactions were carried out in S1 nuclease buffer with 0.5 U of enzyme at 45°C for two hours, after which the reaction was terminated by phenol extraction and ethanol precipitation.

For mRNA termini mapping experiments, S1 nuclease resistant DNA fragments were electrophoresed in either 5% or 8% polyacrylamide/7M urea sequencing gels.[33] An equal amount of polyadenylated RNA isolated in an identical way from normal, non-tumorous callus of *Nicotiana tabacum* var. Havana 425 was used


~3400 2330 2320 2310 2300 2290
AAGCTTCGACAA.....AATATCCGATTATTCTAATAAACGCTCTTTTCTCTTAGGTTTACCCGCCA
HindIII ..23.2.67
Right border of T-DNA
nonT-DNA<-->T-DNA
2280 2270 2260 2250 2240 2230 2220
ATATATCCTGTCAAACACTGATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCATGAGCGGAGAAT
23.2.55N<-
12/14 repeat
2210 2200 2190 2180 2170 2160 2150
TAAGGGAGTCACGTTATGACCCCCGCCGATGACGCGGGACAAGCCGTTTTACGTTTGGAACTGACAGAAC
..23.2.68
..23.2.77N
2140 2130 SstII 2110 2100 2090 2080
CGCAACGTTGAAGGAGCCACTCAGCCGCGGGTTTCTGGAGTTTAATGAGCTAAGCACATACGTCAGAAAC
..2-6 ..23.2.64N 23.2.67<-
mTi40<- ->mTi35
2070 2060 2050 2040 2030 2020 2010
CATTATTGCGCGTTCAAAAGTCGCCTAAGGTCACTATCAGCTAGCAAATATTTCTTGTCAAAAATGCTCC
5' of mRNA
2000 1990 1980 1970 1960 1950 1940
ACTGACGTTCCATAAATTCCCCTCGGTATCCAATTAGAGTCTCATATTCACTCTCAATCCAAATAATCTG
23.22.68<- ..23.2.8
Nopaline synthase gene
1930 1920 1910 1900 1890 1880 1870
CAATGGCAATTACCTTATCCGCAACTTCTTTACCTATTTCCGCCGCAGATCACCATCCGCTTCCCTTGAC
23.2.77N<- 23.2.64N<-
mTi35..
MetAlaIleThrLeuSerAlaThrSerLeuProIleSerAlaAlaAspHisHisProLeuProLeuTh
1860 1850 1840 1830 1820 1810 1800
CGTAGGTGTCCTCGGTTCTGGTCACGCGGGGACTGCATTAGCGGCTTGGTTCGCCTCCCGGCATGTTCCC
->23.11.A1 ->23.11.73 2-6<-
rValGlyValLeuGlySerGlyHisAlaGlyThrAlaLeuAlaAlaTrpPheAlaSerArgHisValPro
1790 1780 1770 1760 1750 1740 1730
ACGGCGCTGTGGGCACCAGCAGATCATCCAGGATCGATCTCAGCAATCAAGGCCAATGAAGGAGTTATCA
..23.2.46N
ThrAlaLeuTrpAlaProAlaAspHisProGlySerIleSerAlaIleLysAlaAsnGluGlyValIleT
1720 1710 1700 1690 1680 1670 1660
CCACCGAGGGAATGATTAACGGTCCATTTAGGGTCTCAGCCTGTGATGACCTTGCCGCAGTTATTCGCTC
..23.2.39N 23.2.8<- ->23.
23.11.A1..
hrThrGluGlyMetIleAsnGlyProPheArgValSerAlaCysAspAspLeuAlaAlaValIleArgSe
1650 1640 1630 1620 1610 1600 1590
CAGCCGTGTACTGATTATTGTAACCCGTGCGGACGTTCACGACAGCTTCGTCAACGAACTCGCCAACTTC
11.A4 23.2.46N<- 23.11.73..
rSerArgValLeuIleIleValThrArgAlaAspValHisAspSerPheValAsnGluLeuAlaAsnPhe


1580 1570 1560 1550 1540 1530 1520
AACGGCGAACTCGCAACAAAGGATATTGTCGTCGTGTGCGGCCATGGCTTCTCCATCAAGTACGAGAGAC
->23.11.A24 ->23.11.A32
AsnGlyGluLeuAlaThrLysAspIleValValValCysGlyHisGlyPheSerIleLysTyrGluArgG
PvuII 1510 1500 1490 1480 1470 1460 1450
AGCTGCGATTCAAGCGAATATTCGAGACGGATAATTCGCCCATAACGTCTAAGCTATCGGATCAAAAAAA
..23.2.68N
23.11.A4.. 23.2.39N<-
lnLeuArgPheLysArgIlePheGluThrAspAsnSerProIleThrSerLysLeuSerAspGlnLysLy
1440 1430 1420 1410 1400 1390 1380
ATGTAACGTCAACATCAAGGAAATGAAAGCGTCTTTCGGACTGTCATGTTTCCCAATTCATCGCGATGAT
23.11.A24..
->23.11.82
sCysAsnValAsnIleLysGluMetLysAlaSerPheGlyLeuSerCysPheProIleHisArgAspAsp
1370 1360 1350 1340 1330 1320 1310
GCTGGCGTGATTGATCTACCCGAAGATACCAAGAACATCTTTGCCCAGCTATTTTCCGCTAGAATCATCT
->23.11.A26 23.11.A32..
AlaGlyValIleAspLeuProGluAspThrLysAsnIlePheAlaGlnLeuPheSerAlaArgIleIleC
1300 1290 1280 1270 1260 1250 1240
GCATCCCGCCGTTGCAAGTGCTATTCTTTTCCAACTATATCACTCATGCGGTTCCGGCAGTCATGAACAT
23.2.68N<-
ysIleProProLeuGlnValLeuPhePheSerAsnTyrIleThrHisAlaValProAlaValMetAsnIl
1230 1220 1210 1200 1190 1180 1170
CGGAAGACTCCGCGACCCAGCCAATTCTCTTACTAAAAGAGCTGAGAAGTGGCTTCTTGAACTAGACGAG
23.11.A26.. ..23.2.>ln
eGlyArgLeuArgAspProAlaAsnSerLeuThrLysArgAlaGluLysTrpLeuLeuGluLeuAspGlu
1160 1150 1140 1130 BamHI 1110 1100
CGAACCCCACGAGCCGAGAAGGGCTTTTTCTTTTATGGTGAAGGATCCAACACTTACGTTTGCAACGTCC
->23.11.86 23.11.82..
->23.11.95
ArgThrProArgAlaGluLysGlyPhePhePheTyrGlyGluGlySerAsnThrTyrValCysAsnValG
1090 1080 1070 1060 1050 1040 1030
AAGAGCAAATAGACCACGAACGCCGGAAGGTTGCCGCAGCGTGTGGATTGCGTCTCAATTCTCTCTTGCA
lnGluGlnIleAspHisGluArgArgLysValAlaAlaAlaCysGlyLeuArgLeuAsnSerLeuLeuGl
1020 1010 1000 990 980 970 960
GGAATGCAATGATGAATATGATACTGACTATGAAACTTTGAGGGAATACTGCCTAGCACCGTCACCTCAT
->23.
->23.11.A22
nGluCysAsnAspGluTyrAspThrAspTyrGluThrLeuArgGluTyrCysLeuAlaProSerProHis
950 940 930 920 910 900 890
AACGTGCATCATGCATGCCCTGACAACATGGAACATCGCTATTTTTCTGAAGAATTATGCTCGTTGGAGG
11.A8 23.2.51n<-
23.11.95..
AsnValHisHisAlaCysProAspAsnMetGluHisArgTyrPheSerGluGluLeuCysSerLeuGluA
880 870 860 850 840 830 820
ATGTCGCGGCAATTGCAGCTATTGCCAAAATCGAAATACCCCTCACGCATGCATTCATCAATATTATTCA
23.11.86.. ->23.11.A2
spValAlaAlaIleAlaAlaIleAlaLysIleGluIleProLeuThrHisAlaPheIleAsnIleIleHi

```
      810        800        790        780        770        760        750
TGCGGGGAAAGGCAAGATTAATCCAACTGGCAAATCATCCAGCGTGATTGGTAACTTCAGTTCCAGCGAC
        ->23.11.A20                          23.11.A8..
  ->23.11.85 23.11.A22..
sAlaGlyLysGlyLysIleAsnProThrGlyLysSerSerSerValIleGlyAsnPheSerSerSerAsp

      740        730        720        710        700        690        680
TTGATTCGTTTTGGTGCTACCCACGTTTTCAATAAGGACGAGATGGTGGAGTAAAGAAGGAGTGCGTCGA
                          ..23.2.2n                                23.11.A20..
LeuIleArgPheGlyAlaThrHisValPheAsnLysAspGluMetValGlu.oc                  ->23

      670        660        650        640        630        620        610
AGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATT
                       Unused                      23.11.A2..
.11.94                 poly A site          23.11.85..

      600        590        580        570        560        550        540
ATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGA

              3' end of mRNA
      530        520        510        500        490        480        470
GATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC

      460        450        440        430        420        410        400
AAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGATCAAACTTCGGTACTGTGTAA
  ->23.11.92               23.2.2n<-

      390        380        370        360        350        340        330
TGACGATGAGCAATCGAGAGGCTGACTAACAAAAGGTATGCCCAAAAACAACCTCTCCAAACTGTTTCGA
                                  23.11.94..          ->23.11.96

      320        310        300        290        280        270        260
ATTGGAAGTTTCTGCTCATGCCGACAGGCATAACTTAGATATTCGCGGGCTATTCCCACTAATTCGTCCT
           23.11.92..

      250        240        230        220        210        200        190
GCTGGTTTGCGCCAAGATAAATCAGTGCATCTCCTTACAAGTTCCTCTGTCTTGTGAAATGAACTGCTGA
      ..23.2.38N

      180        170        160        150        140        130        120
CTGCCCCCCAAGAAAGCCTCCTCATCTCCCAGTTGGCGGCGGCTGATACACCATCGAAAACCCACGTCCG
         ..23.2                          ->23.11.A11
                          23.11.96..->23.11.93

      110        100        90         80         70         60         50
AACACTTGATACATGTGCCTGAGAAATAGGCCTACGTCCAAGAGCAAGTCCTTTCTGTGCTCGTCGGAAA
                                                                23.11.93..

       40         30         20         10   HindIII
TTCCTCTCCTGTCAGACGGTCGTGCGCATGTCTTGCGTTGATGAAGCTT
                                          23.2<-
                                     23.2.38N<-
                                   23.11.A11..
```

as a control in all experiments.

Sizing of nopaline synthase transcripts

Five micrograms of poly A+ RNA isolated from BT37 tumor or normal Havana 425 tissue was glyoxalated and electrophoresed through a 1.8% agarose gel,[38] blotted onto activated APT paper, and hybridized with nick-translated[39] Hind III fragment 23 cloned in pBR325. Hybridization and washing conditions were those used by Alwine et al.[40]

The nopaline synthase transcript was also sized by hybridizing 25 ug of polyadenylated RNA from BT37 tumor to 50 ng of purified, denatured Hind III fragment 23 DNA in 80% formamide, as used for RNA termini mapping experiments. The RNA-DNA hybrids were digested with S1 nuclease and Exonuclease VII and electrophoresed through a 1.8% agarose gel containing 0.03 M NaOH, 0.002 M EDTA, bloted onto activated APT paper, and hybridized to nick-translated Hind III fragment 23 cloned in pBR325.

RESULTS

Nucleotide sequence of the nopaline synthase gene

The kilo-sequencing strategy[27] was used to apply the chain termination method[32] to obtain the nucleotide sequence of the nopaline synthase gene. This strategy is described in detail in the adjacent paper.[27]

Fragment Hind III 23 is an all-bacterial DNA fragment which spans the right border region of the T-DNA of pTi T37. Asymmetric deletions were

Figure 2. DNA sequence from HindIII fragment 23 spanning the right border of T-DNA from the nopaline Ti plasmid pTiT37 of Agrobacterium tumefaciens. This presentation is reversed from the standard map orientation, in order to show the non-template (+) strand of the nopaline synthase gene. The sequence is numbered from the HindIII site on the right, which would be on the left in the standard map orientation. The gap in the sequence between nucleotides 3400 and 2330 is from outside T-DNA, and this sequence is under determined at this time.

The data spans for all of the successful sequencing experiments are shown. The start of each data span is the tail of the arrow (->). The end of each data span (the top of the sequencing gel) is indicated with the name of the experiment with two dots (..). With three exceptions, (2-6, mTi35, mTi40) the data was obtained by the kilo-sequencing(24) strategy, so that each experiment represents a deletion available on M13 transducing phage. Each deletion extends from the tail of the arrow to the end of the HindIII fragment. All strains labelled mTi23.11 carry deletions to the left, with sequencing data toward the right (+ strand sequence on the gel). All strains labelled mTi23.2 have the opposite orientation (deleted toward the right of the arrow, - strand sequence on the gel.)

For example, deletion 23.2.77N has deleted all but 6 codons of the nopaline gene, and retains the promoter region on, therefore, a HindIII-EcoRI fragment. Other labelled features of the sequence are discussed in the text.

generated in the replicative form of M13 phage clones containing Hind III fragment 23 inserted in both orientations at the unique Hind III site of mWB2341. Phage genomes were sized on agarose gels and those that appeared to have deletions differing by 150-200 bp were catalogued and sequenced. In general, this allowed approximately 50 bases of overlap between data from adjacent deletions. The sequence of one strand was confirmed by sequencing appropriate deletions from phage that contained Hind III fragment 23 cloned in the opposite orientation.

The nucleotide sequence from the Hind III site just within T-DNA to a point beyond the right border of T-DNA is shown in Figure 2. The sequence is numbered from 1 to 2400, with nucleotide 1 corresponding to the 5′ adenine residue of the Hind III recognition sequence, and nucleotide 2277 corresponding to the last base pair of T-DNA as previously determined.[5,8,9] An open translation reading frame (1239 base pairs, 413 amino acids) extended from nucleotide 1937 to nucleotide 699, an ochre terminator. Although confirming amino acid sequence data are not available, we identify this as the gene for nopaline synthase, since insertion of a 2.3 Md fragment of the P type plasmid R702, encoding streptomycin and sulphonamide resistance, into the unique Bam HI site at nucleotide 1126 is known to abolish nopaline synthesis.[23] Other reading frames extending 1 kb in either direction from the Bam Hl site at nucleotide 1126 were frequently punctuated with terminator codons. The calculated molecular weight of the polypeptide encoded by this open reading frame is 45460 daltons. Kemp et al[41] estimated the molecular weight of nopaline synthase to be 40,000 daltons by electrophoresis of the purified protein in denaturing acrylamide gels. [If their result were taken as extremely accurate, it would indicate either post-translational processing or a translation start at the second AUG, position 1718 (calculated MW = 38188).]

Localization of the 5′ end of nopaline synthase mRNA

A 920 bp Hpa II fragment extending from nucleotide 1810 within the open reading frame to a point beyond the sequenced region outside T-DNA was labelled at its 5′ extremities with gamma-^{32}P ATP and then cleaved with Sst II to create a fragment of length 314 uniquely labelled at nucleotide 1810. This fragment was hybridized with polyadenylated RNA isolated from BT37 teratoma tissue and then digested with S1 nuclease.

Figure 3 shows an autoradiograph of an 8% sequencing gel of the S1 nuclease-resistant DNA fragments. The autoradiograph reveals a band of radioactive DNA 160 nucleotides long. This indicates that the putative

nopaline synthase mRNA protects a region of DNA extending from the Hpa II site at nucleotide 1810 to approximately nucleotide 1970. A ten-fold increase in S1 nuclease concentration caused only a slight truncation of the protected fragment to one 2 nucleotides smaller.

Reference to an A+G degradation reaction (Maxam and Gilbert, 1980) of the same radiolabelled Hpa II - Sst II fragment run beside the S1 degradation reaction shows the putative 5′ end is located opposite the sequence CTCT, which is AGAG of the + strand, at nucleotides 1974 to 1971. These nucleotides are 25 nucleotides downstream from the sequence CATAAA. This bears a close resemblance to the canonical TATAAA consensus sequence that is generally found

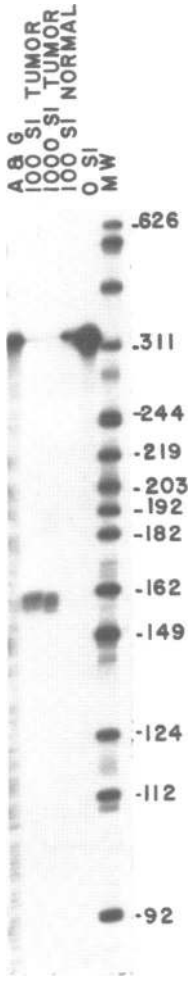


Figure 3. S1 nuclease mapping of the 5′ terminus of the nopaline synthase transcript. Shown is an autoradiogram of an 8% polyacrylamide sequencing gel used to size the S1 nuclease-resistant DNA fragments. A+G: piperidine formate degradation of the 314 nucleotide Hpa II - Sst II fragment. 100 S1 tumor: 100 U/ml S1 nuclease digestion of tumor polyadenylated RNA hybridised to the 314 nucleotide Hpa II - Sst II fragment. 1000 S1 tumor: Similarly, but with ten times more S1/ml. 100 S1 normal: 100 U/ml S1 nuclease digestion of normal callus polyadenylated RNA hybridised to the 314 nucleotide Hpa II - Sst II fragment. 0 S1: A small portion of the intact 314 nucleotide Hpa II - Sst II fragment. MW: 3′ end-labelled Hpa II digestion of pBR322 DNA.

25 to 33 nucleotides upstream of RNA polymerase II transcription initiation sites in eukaryotes.[42] The transcription of T-DNA genes is very sensitive to alpha-amanitin,[43] which implicates RNA polymerase II in their transcription. At position 2051-2043 was found the sequence GGTCACTAT, which (along with 14 other places in the 2400 base pairs shown here) has a 7/9 match with the consensus CCAAT box[44] (GGYCAATCT), often found, as here, 70-80 nucleotides upstream from the proposed site of transcription initiation.

Although a definitive identification of the precise 5′ terminus of the nopaline synthase mRNA awaits a sequence analysis of the mRNA, our S1 nuclease protection data suggest that nopaline synthase transcription initiation signals are consistent with those of several other eukaryotic structural genes.

Localization of the 3′ terminus of nopaline synthase mRNA

A clone of Hind III fragment 23 in pBR325 was cleaved with Bam HI, and labelled at its 3′ termini with the Klenow fragment of E. coli DNA polymerase as described in Materials and Methods. The DNA fragments were then cleaved with Hind III and resolved by electrophoresis in 1% agarose. The fragment containing nucleotides 2 to 1126 was isolated by electroelution, chromatography on DEAE-Sephadex, and ethanol precipitation. The radioactive fragment was hybridized with polyadenylated RNA isolated from BT37 teratoma tissue and then digested with S1 nuclease. Figure 4 shows an autoradiograph of a 5% acrylamide sequencing gel of the S1 nuclease-resistant DNA fragments. It reveals an intense band of 605-610 bases, and two fainter bands of approximately 600 to 598 bases. This result suggests that nopaline synthase mRNA terminates at heterogenous sites, with a major site at nucleotides 514-520, and two minor sites at nucleotides 524 and 526. Inspection of the sequence reveals that the mRNA terminates in AT rich sequences, and as a consequence some of the S1-resistant fragments may be caused by "breathing" of the RNA-DNA duplex during S1 nuclease digestion. Although this was minimized by conducting the S1 digestion at 20 C, it is possible that the two minor, shorter, S1 resistant fragments are artefacts.

These putative 3′ ends are not closely preceded by a perfect canonical polyadenylation signal 5′ AAUAAA, often found 15-20 nucleotides before the point of poly A addition.[45,46] There is however an apparently unused AAUAAA at position 656-651, some 130 nucleotides before the actual 3′ end observed here.

Size of the nopaline synthase mRNA

The distance between the putative 5′ and 3′ termini of the nopaline

synthase gene measures approximately 1460 nucleotides. Two methods were used to measure the size of the mature nopaline synthase transcript and confirm this estimate.

Firstly, polyadenylated RNA isolated from BT37 teratoma tissue was denatured with glyoxal and electrophoresed in 1.8% agarose gels, blotted onto activated APT paper, and hybridized with nick-translated Hind III fragment 23. A radioactive band of 1550-1600 nucleotides was revealed (Figure 5A), which is consistent with the length inferred from sequence data plus a 3′ polyadenylic acid tract, which appears therefore to be 100-150 bases long.

Secondly, 50 ng of purified Hind III fragment 23 was denatured and hybridized with 25 ug of polyadenylated BT37 teratoma RNA in 80% formamide as

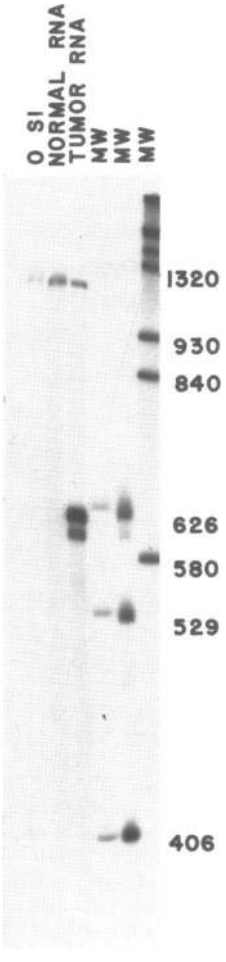


Figure 4. S1 nuclease map of the 3′ terminus of the nopaline synthase transcript. Shown is an autoradiogram of a 5% sequencing gel used to size S1 nuclease-resistant DNA fragments. 0 S1: Intact 1125 nucleotide Hind III - Bam H1 fragment used in the hybridization experiment. Normal RNA: 1125 nucleotide Hind III - Bam H1 fragment hybridized to normal callus polyadenylated RNA and digested with 100 U/ml S1 nuclease. Tumor RNA: 1125 nucleotide Hind III - Bam H1 fragment hybridized to polyadenylated RNA isolated from tumors and digested with 100 U/ml S1 nuclease. MW: Two loading amounts of pBR322 DNA digested with Hpa II and 3′ end-labelled, and 1 loading of phage lambda DNA digested with Eco RI and Hind III and 3′ end-labelled.

described in Materials and Methods, and then digested with either S1 nuclease or Exonuclease VII. The resulting DNA fragments were electrophoresed in an alkaline 1.8% agarose gel together with appropriate molecular weight markers, blotted onto activated APT paper, and hybridized with nick-translated Hind III fragment 23. A radioactive band of 1450-1500 bases was revealed (Figure 5B) when the RNA-DNA duplex was digested with S1 nuclease and with Exonuclease VII. This is consistent with the transcript size of 1460 bases inferred from sequence data, and the observed size on Northern blots of 1550-1600 bases. In addition, the similar size of the Exonuclease VII and S1 nuclease-resistant hybrid molecules shows that RNA protects a continuous length of approximately 1450 nucleotides. This is evidence that the nopaline synthase gene is not

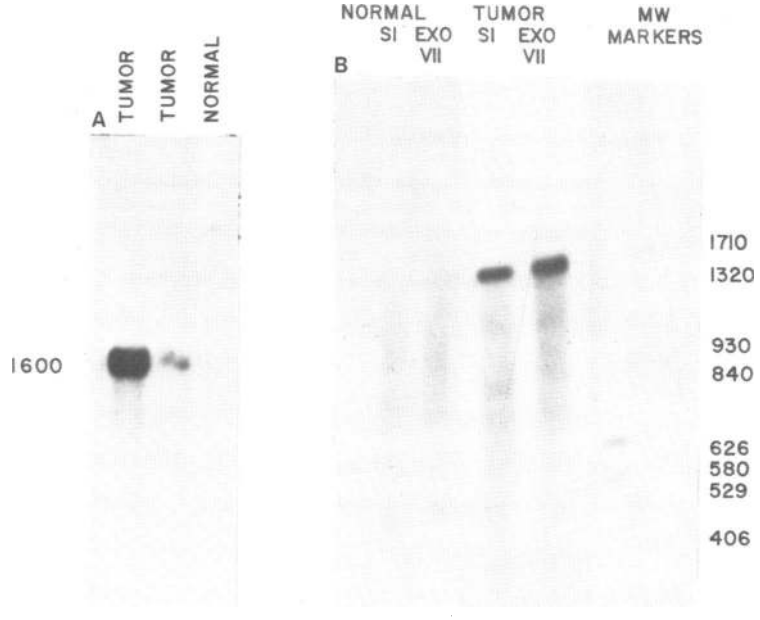


Figure 5. The size of the nopaline synthase transcript.
A) An autoradiogram of a gel of polyadenylated tumor and normal callus RNA blotted onto activated APT paper, and hybridized to nick-translated Hind III fragment 23, is shown. Molecular weight markers (not visible) were glyoxalated plant and E. coli rRNA species, and chimp globin mRNA.
B) An autoradiogram of an alkaline gel of S1 nuclease and Exo VII-resistant DNA fragments blotted onto nitrocellulose and hybridized to nick-translated Hind III fragment 23 is shown. Normal: digestion of normal callus polyadenylated RNA hybridized to purified Hind III fragment 23. Tumor: similarly for tumor polyadenylated RNA. Markers: 3′ end-labelled Hind III and Eco RI digestion of phage lambda DNA, and 3′ end-labelled Hpa II-digested pBR322 DNA.

interrupted by intervening DNA sequences.[47]

DISCUSSION

The data presented show that the gene that encodes nopaline synthase has many similarities to "typical" eukaryotic genes despite its prokaryotic origin. This is not surprising as the gene is expressed in the nucleus of higher plants.

The presence of sequences upstream from the putative 5′ end of the nopaline synthase mRNA that have a high degree of homology to known eukaryotic promotor sequences provides evidence that the putative 5′ end of the mRNA at nucleotide 1964 is the actual 5′ end of the mRNA molecule. Because of this, it is unlikely that plant DNA sequences further upstream from the nopaline synthase gene are involved in the regulation of this gene. The absence of introns in this gene is an unusual feature, as many plant structural genes, like animal genes, contain introns.[48-50]

The nopaline synthase gene promoter is a constitutive plant promoter, as shown by the presence of nopaline in all tissues of shoots that develop from tumors.[51,52] Our determination of the DNA sequence in this region lays the background for engineering of this promoter to express useful proteins in plant cells. Our deletion 23.11.77N, which arose during the sequencing strategy employed, represents a further step in this engineering, since it has deleted all but six codons of the nopaline synthase gene, leaving the nop promoter and the right border of T-DNA on a Hind III - Eco RI fragment. Should deletion of the remaining 6 codons of the nopaline gene be desirable, limited exonuclease digestion could start at the Eco RI site. As it stands in M13, the deletion 77N has fused the sixth codon of nopaline synthase to the sixth codon of the alpha complementing region of beta-galactosidase, in the same reading frame.

ACKNOWLEDGEMENTS

The authors wish to thank Elodee Tuley, Don Bennett and Jim Grichnik for their able assistance. This research was supported by the NIH (grant no. GM24956 to W.M.B.), the American Cancer Society (Faculty Research Award to W.M.B. and grant NP194 to M.D.C.), the Department of Energy (contract AC02-81ER 10889, and Monsanto Corporation (postdoctoral fellowship to M.B.).

*Present address: Plant Breeding Institute, Maris Lane, Trumpington, Cambridge CB2 21Q, UK
§To whom reprint requests should be sent

REFERENCES

1. Zaenen, I., Van Larebeke, N., Teuchy, H., Van Montagu, M. and Schell, J. (1974) Mol.Biol. 86, 109-127.
2. Watson, B., Currier, T.C., Gordon, M.P., Chilton, M.-D. and Nester, E.W. (1975) Bacteriology 123, 255-264.
3. Chilton, M.-D., Drummond, M.H., Merlo, D.J., Sciaky, D., Montoya, A., Gordon, M.P. and Nester, E.W. (1977) Cell 11, 263-271.
4. Yadav, N.S., Postle, K., Saiki, R.K., Thomashow, M.F. and Chilton, M.-D. (1980) Nature (London) 287, 458-461.
5. Zambryski, P., Holsters, M., Kruger, K., Depicker, A., Schell, J., Van Montagu, M. and Goodman, H.M. (1980) Science 209, 1385-1391.
6. Willmitzer, L., De Beuckeleer, M., Lemmers, M., Van Montagu, M. and Schell,J. (1980) Nature (London) 287, 359-361.
7. Simpson, R., O'Hara, P., Kwok, W., Montoya, A.L., Lichtenstein, C., Gordon, M.P. and Nester, E.W., Cell 29 (1982) 1005-1014.
8. Zambryski, P., Depicker, A., Kruger, K., and Goodman, H.M. (1982) J.Mol.Appl.Genet. 1, 361-370.
9. Yadav, N.S., Vanderleyden, J., Bennett, D.R., Barnes, W.M. and Chilton, M.-D., Proc.Nat.Acad.Sci.USA in press.
10. Drummond, M.H., Gordon, M.P., Nester, E.W. and Chilton, M.-D. (1977) Nature (London) 269, 535-536.
11. Gelvin, S.B., Thomashow, M.V., McPherson, J.C., Gordon, M.P., and Nester, E.W. (1982) Proc. Natl. Acad. Sci. USA 79, 76-80.
12. Willmitzer, L., Simons, G., and Schell, J. (1982) The EMBO Journal 1, 139-146.
13. McPherson, J.C., Nester, E.W. and Gordon, M.P. (1980) Proc. Natl. Acad. Sci. USA 77, 2666-2670.
14. White, P.R. and Braun, A.C. (1942) Cancer Res. 2, 597-617.
15. Petit, A., Delhaye, S., Tempe, J. and Morel, G. (1970) Physiol. Veg. 8, 205-213.
16. Petit. A. and Tempe, J. (1978) Molec.Gen.Genet. 167 (167) 147-155.
17. Schroder, J. and Schroder, G. (1981) FEBS Letters 129, 166-68.
18. Murai, N. and Kemp, J. (1982) Proc. Natl. Acad. Sci. USA 79, 86-90.
19. Garfinkel, D.J., Simpson, R.B., Ream, L.W., White, F.F., Gordon, M.P., and Nester, E.W. (1981) Cell 27, 143-153.
20. 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.
21. Gelvin, S.B., Gordon, M.P., Nester, E.W. and Aronson, A.A. (1981) Plasmid 6, 17-29.
22. Hernalsteens, J.P., Van Vliet, F., De Beuckeleer, M., Depicker, A., Engler, G., Lemmers, M., Holsters, M., Van Montagu, M. and Schell, J. (1980) Nature (London) 287, 654-656.
23. Leemans, J., Shaw, C., Deblaere, R., De Greve, H., Hernalsteens, J.P., Maes, M., Van Montagu, M. and Schell, J. (1981) J. Mol. Appl. Genet. 1, 149-164.
24. Bolivar, F. (1978) Gene 4, 121-136.
25. Prentki, P., Karch, F., Iida, S. and Meyer, J. (1981) Gene 14, 289-299.
26. Williams, B.C. and Blattner, F.R. (1980) in Genetic Engineering (Setlow, J.K. and Hollaender, A., eds.) vol 2, 201-283.
27. Barnes, W.M. and Bevan, M.(1982) Nucleic Acid Res.
28. Barnes, W.M. (1977) Science 195, 393-394.
29. Blair, D.G., Sherratt, D.J., Clewell, D.B. and Helinski, D.R. (1972) Proc. Natl. Acad. Sci. USA 69, 2518-2522.
30. Klenow, H., Overgaard-Hansen, K. and Patkar, S.A. (1971) Eur. J. Biochem. 22, 371-381.
31. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-323.

32. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
33. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
34. Anderson, S.(1981) Nucleic Acids Res. 9, 3015-3027.
35. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology (Gross, L. and Moldave, K, eds.) vol 65, 499-560.
36. Chirgwin, J.M., Przybyla, A.E., Macdonald, R.J., and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
37. Bantle, J.A., Holm, I.H. and Hahn, W.E. (1978) Anal. Biochemistry 72, 413-427.
38. McMaster, G.F. and Carmichael, G.G. (1977) Proc. Natl. Acad. Sci. USA 74, 4835-4838.
39. Maniatis, T., Jeffrey, A., and Kleid, D.G. (1975) Proc. Natl. Acad. Sci. USA 72, 1184-1188.
40. Alwine, J.C., Kemp, D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R. and Wahl, G.M. (1979) Methods in Enzymology (ed. by R. Wu) vol 68, 220-242.
41. Kemp, J.D., Sutton, D.W. and Hack, E. (1979) Biochemistry 18, 3755-3760.
42. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochemistry 50, 349-383.
43. Efstradiatis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., and Proudfoot, J.J. (1980) Cell 21, 653-668.
44. Willmitzer, L., Schmalenbach, W. and Schell, J. (1981) Nucleic Acid Res. 9, 4801-4812.
45. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
46. Benoist, R., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acid Res 8, 127-142.
47. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
48. Shah, D.M., Hightower, R.C. and Meagher, R.B. (1982) Proc.Natl.Acad.Sci. USA 79, 1022-1026.
49. Sun, S.M., Slightom, J.L., and Hall, T.C. (1981) Nature 289 37-41.
50. Wiborg, O., Hyldig-Nielsen, J.J., Jensen, E.O., Paludan, K. and Marcker, K.A. (1982) Nucleic Acids Res 10, 3487-3494.
51. Wood, H.H., Binns, A.N. and Broun, A.C. (1978) Differentiation 11, 175-180.
52. Wullems, G.J., Molendijk, L., Ooms, G. and Schilperoort, R.A. (1981) Cell 24, 719-727.