Mol Gen Genet (1990) 221:256–266

MGG

The *virB* operon of *Agrobacterium tumefaciens* pTiC58 encodes 11 open reading frames

Gretchen A. Kuldau[1], Guido De Vos[1], John Owen[2], Gretchen McCaffrey[3], and Patricia Zambryski[1]

[1] Department of Plant Biology, University of California, Berkeley, CA 94720, USA
[2] Department of Biochemistry, Royal Holloway Hall and Bedford New College, Egham Hill, Egham, Surrey TW20 OEX, UK
[3] Department of Pharmacology, University of San Francisco, San Francisco, CA 94143, USA

Summary. *Agrobacterium tumefaciens* genetically transforms plant cells by transferring a copy of its T-DNA to the plant where it is integrated and stably maintained. In the presence of wounded plant cells this process is activated and mediated by the products of the *vir* genes which are grouped into six distinct loci. The largest is the *virB* locus spanning 9.5 kb. Transposon mutagenesis studies have shown that *virB* gene products are required for virulence but their functions remain largely unknown. To provide information relevant to understanding the function of VirB polypeptides, the nucleotide sequence of the *virB* operon from a nopaline plasmid, pTiC58, is presented here. Eleven open reading frames (ORFs) are predicted from this sequence. The predicted sizes of 10 of the 11 VirB polypeptides are verified by specific expression in *Escherichia coli*. Only the product of the smallest ORF potentially encoding a 5.8 kDa polypeptide has not been detected. The initiation of translation of five *virB* ORFs occurs at codons that overlap the termination codons of the ORF immediately upstream; thus, translational coupling may be an important mechanism for efficient translation of the large *virB* polycistronic mRNA. Based on hydropathy plot analysis nine of the *virB* ORFs encode proteins that may interact with membranes; these data support the earlier hypothesis (Engstromm et al. 1987) that *virB* gene products may form a membrane pore or channel to mediate exit of the T-DNA copy (T-strands) from *Agrobacterium* into the plant cell. A comparison of the two published octopine *virB* sequences with the nopaline sequence presented here is made.

Key words: *Agrobacterium* – Vir proteins – *virB* DNA sequence – Membrane proteins – T-DNA transport

Introduction

Interest in *Agrobacterium tumefaciens* derives from its ability to transform plant cells. This gram-negative soil bacterium causes crown gall disease by transferring a copy of its T-DNA (transferred DNA) from its Ti(tumor inducing) plasmid to susceptible plant cells where it is stably integrated and maintained (reviewed in Zambryski 1988a, b). While the T-DNA codes for plant growth regulators which cause undifferentiated plant cell growth, a separate region of the large (>200 kb) Ti plasmid, the *vir* region, encodes products to mediate T-DNA movement. Six *vir* complementation groups have been identified. Four of these (*virA, virB, virD* and *virG*) are essential for transformation, and two (*virE* and *virC*) are only required on certain hosts (Stachel and Nester 1986). Transcription of *vir* genes only occurs in the presence of wounded plant cells which secrete small phenolic compounds such as acetosyringone (Stachel et al. 1985). Plant phenolics act in concert with the *virA* and *virG* gene products to regulate *vir* expression (Stachel and Zambryski 1986). VirA and VirG are thought to act analogously to other bacterial two component systems: VirA is a membrane sensor that detects plant signals and transmits this information to intracellular VirG to direct specific interaction with *vir* promoters to activate *vir* transcription (Ronson et al. 1987).

After induction of *vir* gene expression, molecular reactions occur to generate a transferrable T-DNA copy. Sequence specific single-stranded nicks can be detected at the 25 bp T-DNA borders and single-stranded copies of the bottom strand of the T-DNA (T-strands) are found in *Agrobacterium* cells (Stachel et al. 1986). The endonucleolytic activity at the T-DNA borders and the production of T-strands are dependent on the *virD1* and *virD2* gene products (Yanofsky et al. 1986; Stachel et al. 1987; Veluthambi et al. 1987). T-strands are likely associated with Vir proteins, forming a T-complex (Citovsky et al. 1988a). Specifically, the 5′ end of the T-strand is tightly associated with the VirD2 protein (Ward and Barnes 1988; Herrella-Estrella et al. 1988; Howard et al. 1989) and the length of the molecule is coated with a single-stranded DNA binding protein which has been identified as the *virE2* gene product (Gietl et al. 1987; Das 1988; Christie et al. 1988; Citovsky et al. 1988a, 1988b).

While much is known about the early events in the *Agrobacterium* mediated plant cell transformation process, little is known about the molecular mechanism(s) of the actual transfer of T-complexes from the bacterial cell to the plant cell. Potentially a membrane channel or pore may facilitate movement of the T-complex out of the bacterial cell and into the plant cell. This T-complex must presumably pass through the bacterial inner membrane, peptidoglycan layer, and outer membrane, then through the plant cell wall, cytoplasmic membrane and finally into the plant cell nucleus. Previous studies have shown that *virB* is not directly involved in T-DNA processing (e.g. Stachel et al. 1987) and that *virB* protein products are localized to the bacterial

Offprint requests to: P. Zambryski

envelope (Engstromm et al. 1987). Membrane localized VirB proteins could comprise the channel for T-complex export.

While there have been two published reports of the *virB* sequence from an octopine strain (Ward et al. 1988; Thompson et al. 1988), our laboratory is interested in studying the molecular mechanism(s) of the transformation process directed by nopaline type strains. Here we present the complete sequence of the *virB* operon from pTiC58 and verify the predicted sizes of the open reading frames (ORFs) by specific expression of the individual *virB* protein products. The results show that the octopine and nopaline *virB* regions are highly homologous in overall nucleic acid and amino acid sequence. While the sizes of the VirB proteins predicted by our sequence data and the sizes of the VirB proteins identified earlier by Engstromm et al. (1987) agree with each other, they do not agree with some of the predictions from the octopine sequence data of Thompson et al. (1988) and Ward et al. (1988). These present results clarify existing confusion regarding sizes of VirB proteins and indicate several possible errors in the published nucleic acid sequences of the octopine *virB* region.

Materials and methods

General methods and materials. General molecular techniques were performed as described by Maniatis et al. (1982) except were noted. DNA restriction enzymes, Klenow fragment and DNA ligase were from New England Biolabs, Sequenase was from United States Biochemical (USB), *Asp*718 and S1 nuclease were from Boehringer Mannheim, and exonuclease III was from Bethesda Research Laboratories. Nucleotides used for sequencing with Klenow enzyme were from Pharmacia. Translabel, a high specific activity mixture of [^{35}S]methionine and [^{35}S]cysteine was obtained from ICN.

Sequencing. A series of overlapping deletions was made for both strands of the nopaline *virB* region using six subclones of the *Hin*dIII fragments shown in Fig. 1. The upstream *Hin*dIII fragment extends 4.4 kb farther to the left than shown in Fig. 1. Each of the three *Hin*dIII fragments from pGVO310 (Depicker et al. 1980) was subcloned into pUC119 (Vieira and Messing 1987) in both orientations and deletions were made using exonuclease III and nuclease S1. In addition to the short overlapping deletions shown in Fig. 1, subclones were made using available restriction sites to confirm some areas of the sequence, for example, to sequence across the upstream *Hin*dIII site. Four oligonucleotides complementary to *virB* were also used for unclear sections; one was used to sequence across the downstream *Hin*dIII site. Strain XLIBlue (Stratagene) was used to maintain deletions and to generate single-stranded templates with helper phage M13KO7 (Vieira and Messing 1987). Occasionally, double-stranded (ds) templates were prepared by an alkaline lysis method (Maniatis et al. 1982) followed by 30 min incubation with RNase A and a polyethylene glycol precipitation. The ds templates were annealed to primers according to Chen and Seeburg (1985) and sequenced using Sequenase. Dideoxy chain termination sequencing was performed using either the Klenow fragment of DNA polymerase I or Sequenase as described by the manufacturer. Computer sequence analysis programs were from Intelligenetics, Hugo Martinez at the Biomathematics Computation Laboratory, Department of Biochemistry and Biophysics, UCSF, and Guido de Vos (unpublished).

Subcloning. Plasmid pGK215 is an exonuclease III generated deletion of the proximal *Hin*dIII fragment (Figs. 1, 3A) which contains *virB* sequences starting 125 bp upstream of the transcription start site (Das et al. 1986). This construct (pGK215) was used to generate pGK216 by digestion with *Hin*dIII followed by ligation, in the correct orientation for *virB* transcription, with the 4.0 kb *Hin*dIII fragment from pGVO310. Plasmid pGK217, containing all of the *virB* operon was constructed by ligating, again in the correct orientation for *virB* transcription, the distal *Hin*dIII fragment of pGVO310 into partially digested pGK216 at its second *Hin*dIII site. A premature translation stop in VirB11 made made by digestion of pGK217 with *Asp*718 at the single site in VirB11, using Klenow enzyme to form blunt ends, and religating to generate pGK218. A subclone, pGK219, expressing VirB1, VirB2, VirB3 and the first 84 amino acids of VirB4 was made by ligating *virB* sequences from an exonuclease III deletion in pUC119 into the expression vector T7-7 (gift from F. Studier). The *virB* sequences were removed from pUC119 with *Eco*RI and *Hin*dIII and ligated into the *Eco*RI and *Hin*dIII sites of T7-7. The regenerated *Eco*RI site in this new construct was then cut with *Eco*RI and filled in using Klenow enzyme. This new construct has an in-frame fusion of 18 amino acids to VirB1. Plasmid pYT910, expressing VirB9 and VirB10 was made from an exonuclease III deletion in pUC119 which begins just inside the *virB9* coding region. The *virB* sequences were removed from pUC119 as a 1.95 kb *Eco*RI fragment and ligated into the *Eco*RI site of pT7-7 in the correct orientation for transcription. The resulting fusion to VirB9 is predicted to be missing its first eight amino acids and to have gained four amino acids from pT7-7 for a net loss of four from the size or original protein. This construct also contains the entire *virB10* coding sequence as the distal *Eco*RI site begins 1 bp after the TAA translation stop of *virB10*.

Expression of VirB proteins. T7 expression plasmids were maintained and induced (Studier and Moffat 1986) in BL21DE3(pLysE) cells (gift from F. Studier). Cells grown to 0.5–1.0 OD_{600} were induced with 0.4 μM isopropyl-thio-β-galactoside (IPTG) for 1 h and then treated with 200 μg/ml rifampicin for 1 h before harvesting. After 1 to 3 h of induction, cells were labeled with 5 μCi/ml of Translabel for 10 min and harvested by centrifugation. Maxi cells (strain RB901) were prepared by a modification of the method of Sancar et al. (1981). Cells (5 ml; $OD_{600}=0.5$–1.0) were irradiated for 40 s in plastic petri dishes at 360 ergs/mm^2 per 10 s with swirling every 10 s. After 1 h of growth in Hershey's sulfate-free medium, the cells were labeled with 5 μCi/ml of [^{35}S]methionine/cysteine for 30 min and harvested by centrifugation. Labeled cells were resuspended in Laemmli loading buffer and separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli 1970). Gels were exposed directly to X-ray film.

Results and discussion

Nucleotide sequence of the nopaline virB *region*

A restriction map of the *virB* region of the nopaline pTiC58 plasmid is shown in Fig. 1. The *virB* transcriptional unit

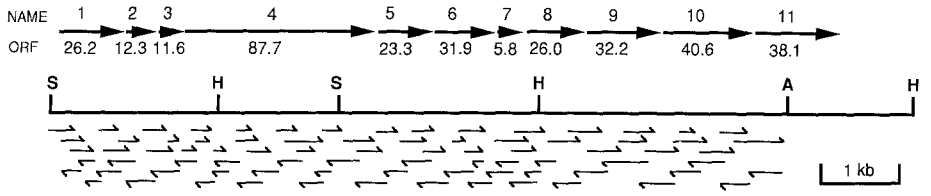

Fig. 1. Predicted open reading frames (ORFs) and sequencing strategy of the *virB* operon. A restriction map of the nopaline pTiC58 *virB* region is shown. *Large arrows* above the restriction map indicate ORFs. The ORF number is given above these arrows and below is shown the predicted molecular weight of the ORF in kDa. *Small arrows* indicate subclones used for sequencing this region. S, *Sac*I; H, *Hin*dIII; A, *Asp*718

of the octopine Ti plasmid A6 was first defined by Stachel and Nester (1986) using transposon mutagenesis. A heteroduplex map between octopine and nopaline *vir* regions (Engler et al. 1981) showed a region in pTiC58 homologous to the octopine *virB* region. Based on this information, we chose to sequence the area beginning at the *Sac*I site shown in Fig. 1 through to the distal *Asp*718 site. The nucleotide sequence of both strands of the *virB* operon from pTiC58, encompassing 9072 bp, was determined. Figure 1 shows the sequencing strategy and the sizes of the predicted ORFs. The nucleotide sequence and predicted amino acid sequence of the open reading frames are shown in Fig. 2. The final 758 nucleotides of the operon shown in Fig. 2 are from a previous report (Powell et al. 1987). *virB* codes for 11 proteins, designated VirB1 through VirB11.

The limits of each of the 11 VirB polypeptides were deduced from the nucleotide sequence of the *virB* region. Each polypeptide corresponds to a unique contiguous sequence of amino acids that initiates with a methionine codon and ends with a termination codon. Although rare, it is possible that other initiation codons are used; protein sequencing or detailed genetic studies are required to establish the exact starts of translation. However, the assignments of methionine initiation codons made here are consistent with several other criteria important for translation of mRNA into protein. First, the initiation codon should be preceded by a sequence that shows homology to the 3′ end of ribosomal 16S RNA, the so-called Shine-Dalgarno (SD) sequence. The SD sequences proposed for each ORF are derived from homologies to both *Agrobacterium* (Yang et al. 1985) and *Escherichia coli* (Shine and Dalgarno 1974) 16S ribosomal RNAs. Interestingly, while there is a slight difference in the exact SD sequence from each bacterium, the SD sequences for the VirB ORFs appear to be equally suited for *Agrobacterium* or *E. coli* ribosomes. Second, the initiation codon should be properly spaced with respect to the SD sequence, usually 5 to 13 nucleotides between the end of the SD and the first base of the ATG methionine codon. Third, the possibility for translational coupling has to be taken into account. When downstream ATG start codons are near (<3 bp) or are overlapped by the immediately upstream stop codon in polycistronic messages, the translation of the downstream protein is dependent on translation of the immediately upstream protein (reviewed in Gold 1988). This was first described for the *trp* operon of *E. coli* and has been called translational coupling (Oppenheim and Yanofsky 1980). Such translational coupling is significantly reduced in the absence of an SD sequence (Das and Yanofsky 1984). Translational coupling is assessed here since it is noteworthy that, as described below, five of the VirB polypeptides likely use this mechanism to initiate their synthesis. Fourth, the deduced initiation sites used should delimit ORFs of sizes that are confirmed by studies actually to express and detect the individual protein products. The results in Fig. 3 are from experiments specifically to assay the sizes of individual VirB proteins following their expression and analysis by gel electrophoresis.

Table 1 summarizes our evaluation of each of the predicted VirB polypeptides for the four criteria mentioned above. Most of the predicted ORFs fulfill at least three of the four criteria evaluated. VirB2, VirB3, VirB4, VirB9, and VirB10 may fulfill the fourth criterion, possibly utilizing translational coupling to initiate their synthesis. Genetic studies (Stachel and Nester 1986) have provided evidence that the region encompassing the *virB* complementation group most likely is transcribed as a single mRNA greater than 9000 nucleotides long. Translational coupling may help keep ribosomes bound to this long mRNA minimizing exposure of intercistronic sequences to nucleolytic attack.

While the assignment of the start of VirB10 appears to fulfill three out of the four criteria tested in Table 1, expression studies show that the ORF in this region encodes a polypeptide with a mobility of 51 kDa on SDS-PAGE (Fig. 3B, lanes 2, 5, 6, 11). In contrast, we predicted a 40.6 kDa polypeptide from the nucleotide sequence data. Since our assignment of the start of VirB10 is potentially in error, we have indicated another possible start site on the nucleotide sequence shown in Fig. 2, designated 10a. However, this upstream start does not have a good SD site, nor does it allow translational coupling. Thus, we favor the assignment of the ATG start that would encode the 40.6 kDa polypeptide. We suggest that this protein has an anomolous mobility on SDS-PAGE; this hypothesis is supported by the fact that the VirB10 polypeptide appears to migrate as a diffuse rather than a sharp protein band.

VirB6, VirB7, and VirB8 do not fulfill our criteria for translational initiation. There are two possible initiation (ATG) start sites upstream of the VirB6 ORF, indicated 6a and 6b on Fig. 2. Neither site has a good SD sequence, and neither allows translational coupling. The use of the 6a start, within the coding sequence of *virB5* is unlikely, since there is a low probability for a ribosome to initiate internal to sequences undergoing active translation; note that the VirB5 product is abundantly produced, suggesting that the 6a start does not interfere with translation of VirB5 (Fig. 3B, lanes 2, 4, 5). Ribosome drifting has been proposed to explain reinitiation in the proper frame at sites near frame shift mutations (Shinedling et al. 1987); however, the probability that the ribosome could drift 24 codons backward after terminating VirB5 seems low. Thus, we suggest that the start site for VirB6 is at the 6b site. The lack of a good SD sequence upstream of the 6b start

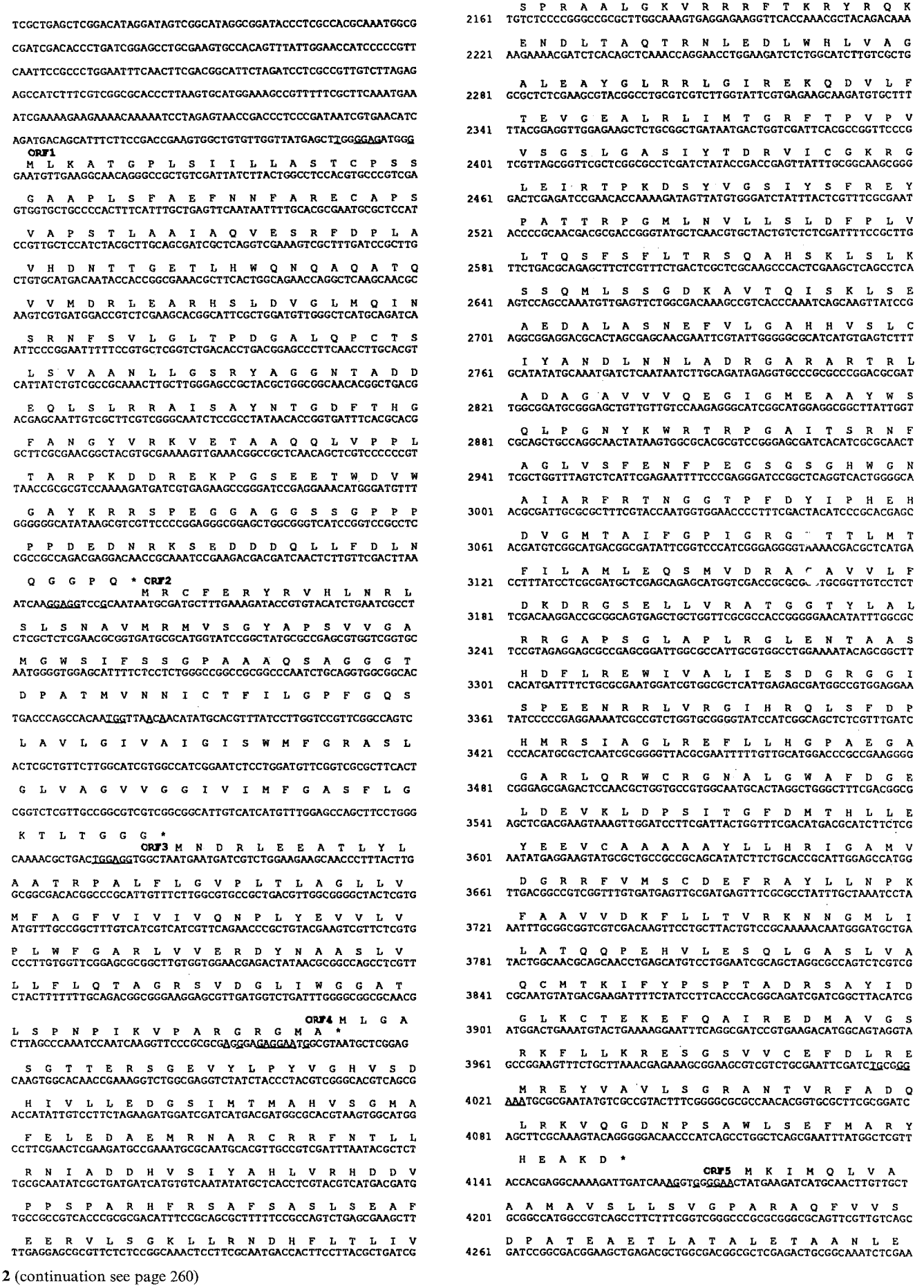

Fig. 2 (continuation see page 260)

```
      Q  T  I  T  M  V  A  M  L  T  S  A  Y  G  V  T  G  L  L  T
4321  CAGACCATAACGATGGTGGCGATGTTAACCTCGGCTTATGGCGTCACCGGCCTACTAACT

      S  L  N  Q  K  N  Q  Y  P  S  T  R  D  L  D  T  E  M  F  S
4381  TCGCTCAACCAAAAAAATCAGTATCCCTCGACCAGGGACTTGGACACGGAAATGTTTTCG

      P  R  M  P  M  S  T  T  A  R  A  I  T  T  D  T  D  R  A  V
4441  CCGCGAATGCCAATGTCGACCACGGCACGTGCGATCACCACCGATACAGATCGCGCCGTA

      V  G  G  D  A  E  A  D  L  L  R  S  Q  I  T  G  S  A  N  S
4501  GTTGGTGGCGACGCTGAAGCGGATCTGTTGCGATCGCAGATCACCGGTTCCGCAAATAGC

      A  G  I  A  A  D  N  L  E  T  M  D  K  R  L  T  A  N  A  E
4561  GCGGGCATTGCGGCTGACAACCTGGAGACGATGGACAAACGCTTAACAGCGAATGCCGAG

      T  S  T  Q  L  S  R  S  R  N  I  M  Q  A  T  V  T  N  G  L
4621  ACCTCGACACAGCTTTCTCGCTCTCGCAATATCATGCAGGCAACCGTTACCAACGGTTTG

      L  L  K  Q  I  H  D  A  M  I  Q  N  V  Q  A  T  S  L  L  T
4681  CTTCTCAAGCAGATCCATGACGCAATGATTCAAAATGTACAGGCGACCAGCTTGTTAACG

                    ORF6a M  R  R  K  R  R  P  L  N  V  R  S
      M  T  T  A  Q  A  G  L  H  E  A  E  E  A  A  A  Q  R  K  E
4741  ATGACCACCGCGCAGGCTGGCCTTCATGAGGCGGAAGAGGCGGCCGCTCAACGTAAGGAG

      I  K  R  L  R  S  S  L  G  P  S  P  E  T  G  R  L  I  R  A
      H  Q  K  T  A  V  I  F  G  A  V  P  *
4801  CATCAAAAGACTGCGGTCATCTTTGGGGCCGTCCCCTGAGACTGGGCGACTTATTCGTGC

      A  H  L  H  I  S  G  C  E  F  I  V  S  D  I  L  R  Q  A  S
4861  CGCCCATCTGCACATTTCAGGATGCGAGTTCATTGTATCCGACATTCTGCGACAAGCCAG
                    ORF6b
      Q  G  Q  V  Q  T  M  N  F  T  I  P  A  P  F  T  A  I  H  T
4921  TCAAGGTCAGGTCCAAACGATGAATTTCACGATCCCGGCGCCGTTTACGGCCATTCATAC

      I  F  D  L  A  F  T  T  S  L  D  T  M  L  G  T  I  Q  E  A
4981  GATCTTCGATCTAGCCTTTACGACAAGCCTGGACACAATGCTTGGGACGATCCAAGAGGC

      V  S  A  P  L  V  A  C  V  T  L  W  I  I  V  Q  G  I  L  V
5041  GGTGAGCGCGCCATTGGTCGCCTGCGTCACTCTTTGGATTATCGTTCAGGGTATTCTGGT

      M  R  G  E  I  D  T  R  G  G  I  T  R  V  I  T  V  T  V  V
5101  CATGCGTGGCGAAATCGACACGCGCGGCGGTATCACTCGGGTGATCACGGTCACCGTCGT

      V  A  L  V  V  G  Q  A  N  Y  H  D  Y  V  V  S  V  F  E  E
5161  TGTTGCCCTTGTCGTCGGGCAGGCCAACTACCACGACTATGTGGTTTCAGTCTTTGAAGA

      T  I  P  N  F  I  Q  Q  F  S  G  S  G  L  P  L  Q  T  I  P
5221  GACGATTCCAAACTTTATTCAGCAGTTTAGTGGCAGCGGCCTGCCTCTGCAGACCATTCC

      A  Q  L  D  T  M  F  A  L  T  Q  A  A  F  Q  R  I  A  S  E
5281  CGCTCAGCTCGATACAATGTTCGCCCTAACCCAGGCTGCATTTCAGAGAATTGCATCTGA

      I  G  P  M  N  D  Q  D  I  L  A  F  Q  G  A  Q  W  V  F  Y
5341  AATCGGCCCGATGAATGACCAGGACATCCTTGCTTTCCAGGGGGCTCAGTGGGTCTTTTA

      G  T  L  W  S  A  F  G  I  Y  D  A  V  G  I  L  T  K  V  L
5401  CGGCACGCTCTGGTCTGCCTTCGGAATCTACGACGCCGTCGGAATTCTCACGAAAGTGCT

      L  A  I  G  P  L  I  L  T  G  Y  I  F  D  R  T  R  D  I  A
5461  TCTGGCGATCGGCCCTTTGATCCTCACGGGATATATCTTTGATCGCACGCGCGACATCGC

      A  K  W  I  G  Q  L  I  T  Y  G  L  L  L  L  L  L  N  L  V
5521  GGCAAAGTGGATCGGGCAACTTATCACCTATGGTCTCCTACTTCTCCTGCTAAACCTCGT

      A  T  I  V  I  L  T  E  A  T  A  L  T  L  M  L  G  V  I  T
5581  GGCGACAATAGTCATCCTAACCGAAGCGACTGCACTCACGCTTATGCTTGGTGTAATCAC

      L  A  G  T  T  A  A  K  I  I  G  L  Y  E  L  D  M  F  F  L
5641  CCTTGCCGGTACGACCGCGGCCAAGATCATTGGTCTTTACGAACTCGACATGTTCTTTCT

      T  G  D  A  L  I  V  A  L  P  A  I  A  G  N  I  G  G  S  Y
5701  GACTGGTGACGCACTCATTGTCGCTTTGCCGGCAATCGCCGGCAACATCGGAGGCAGTTA

      W  S  G  A  T  Q  S  A  N  S  L  Y  R  R  F  A  Q  V  D  R
5761  TTGGAGCGGCGCGACCCAATCTGCCAACAGCTTGTACCGTCGCTTCGCTCAGGTCGATCG

      R  *                          ORF7 M  K  Y  C  L  L
5821  TCGTTAGGTCCCGCAAAAAAATTCACGTCGGTGGAGAACTCTATGAAATATTGCCTGCTG

      C  L  A  L  A  L  G  G  C  Q  T  N  D  K  L  A  S  C  K  G
5881  TGCCTGGCTCTCGCTTTAGGCGGCTGCCAGACAAACGATAAATTGGCGAGCTGCAAAGGC

      P  I  F  P  L  N  V  G  R  W  Q  P  T  P  S  D  L  Q  L  S
5941  CCAATATTCCCGCTGAATGTGGGGCGATGGCAGCCTACGCCGTCAGATCTTCAGCTCAGC

                ORF8 M  K  G  S  E  Y  A  L  L  V  A  R  E  T  L
      N  V  G  G  R  H  E  G  V  *
6001  AACGTAGGTGGTCGCCATGAAGGGGTCTGAATACGCCTTGCTAGTAGCGCGGGAAACCCT

      A  E  H  Y  K  E  A  F  Q  T  A  R  A  K  S  A  R  R  L  S
6061  GGCTGAGCACTACAAGGAAGCTTTCCAAACTGCGCGTGCGAAATCAGCGCGGCGCCTCTC

      K  V  I  A  A  V  A  T  I  A  V  L  G  N  V  A  Q  A  F  T
6121  CAAGGTCATTGCAGCTGTCGCAACCATCGCGGTTTTGGGGAATGTTGCGCAAGCCTTCAC

      I  A  T  M  V  P  L  I  R  L  V  P  V  Y  L  W  I  R  P  D
6181  AATTGCCACCATGGTGCCGCTGATCAGGCTTGTGCCGGTATATCTTTGGATACGGCCGGA

      G  T  V  D  S  E  V  S  V  S  R  L  P  A  T  Q  E  E  A  V
6241  TGGCACCGTTGACAGCGAGGTGTCCGTCTCCCGATTGCCTGCAACTCAAGAGGAGGCCGT

      V  N  A  S  L  W  E  Y  V  R  L  R  E  S  Y  D  A  D  T  A
6301  CGTTAACGCCTCACTGTGGGAGTATGTTCGGCTGCGCGAGAGCTATGATGCCGATACCGC

      Q  Y  A  Y  D  L  V  S  N  F  S  A  P  M  V  R  Q  N  Y  Q
6361  CCAGTATGCCTATGACCTGGTCTCGAACTTCAGCGCCCCAATGGTGCGCCAAAATTATCA

      Q  F  F  N  Y  P  N  P  T  S  P  Q  V  I  L  G  K  H  G  R
6421  GCAATTCTTCAATTATCCCAATCCAACTTCGCCTCAAGTCATCCTCGGCAAACACGGCAG

      L  E  V  E  H  I  A  S  N  D  V  T  P  G  V  Q  Q  I  R  Y
6481  GCTAGAGGTCGAACACATCGCTTCGAATGATGTTACTCCGGGTGTGCAGCAAATTCGCTA

      K  R  T  L  I  V  D  G  K  M  P  M  A  S  T  W  T  A  T  V
6541  TAAGCGAACCCTCATCGTTGACGGCAAAATGCCGATGGCGAGCACTTGGACTGCTACGGT

      R  Y  E  K  V  T  S  L  P  G  R  L  R  L  T  N  P  G  G  L
6601  TCGTTATGAAAAGGTGACCAGCTTGCCCGGCAGATTGAGACTGACCAACCCGGGAGGCTT

      V  V  T  S  Y  Q  T  S  E  D  T  V  S  N  A  G  H  S  E  P
6661  GGTTGTCACCTCCTACCAGACATCGGAAGATACCGTTTCGAACGCAGGCCACAGCGAACC

      * ORF9
      M  T  K  K  A  F  L  T  L  A  C  L  L  F  A  A  I  G  A  R
6721  ATGACGAAAAAAGCATTTCTCACTCTGGCATGTTTACTTTTTGCGGCGATTGGCGCGAGG

      A  E  D  T  P  T  A  G  R  L  D  P  R  M  R  Y  L  A  Y  N
6781  GCTGAAGACACGCCAACGGCGGGCAGACTTGATCCGCGCATGCGTTATCTCGCTTACAAT

      P  D  Q  V  V  R  L  S  T  A  V  G  A  T  L  V  V  T  F  G
6841  CCCGATCAAGTGGTGCGCCTTTCAACGGCGGTTGGAGCCACTTTGGTTGTTACTTTCGGG

      A  N  E  T  V  T  A  V  A  V  S  N  S  K  D  L  A  A  L  P
6901  GCTAACGAAACGGTGACAGCTGTTGCCGTTTCCAATAGCAAAGATCTCGCGGCCCTTCCA

      R  G  N  Y  L  F  F  K  A  S  K  V  L  P  P  Q  P  V  V  V
6961  CGCGGAAATTATCTTTTCTTCAAGGCTAGCAAGGTTCTCCCACCCCAGCCAGTGGTCGTG

      L  T  A  S  D  A  G  M  R  R  Y  V  F  S  I  S  S  K  T  L
7021  CTAACTGCGAGTGACGCCGGTATGCGACGCTACGTTTTCAGCATCTCTTCCAAGACGCTG

      P  H  L  D  K  E  Q  A  D  L  Y  Y  S  V  Q  F  A  Y  P  A
7081  CCGCACCTCGATAAAGAGCAGGCCGATCTCTACTATAGCGTACAATTCGCTTACCCTGCC

      D  D  A  A  A  R  Q  K  A  A  Q  E  K  A  V  A  D  R  I  R
7141  GATGACGCAGCGGCTCGCCAGAAGGCGGCACAAGAGAAGGCTGTTGCAGACCGTATACGT

      A  E  A  Q  Y  Q  Q  R  A  E  G  L  L  E  Q  P  A  T  T  V
7201  GCGGAAGCGCAATATCAACAGAGAGCAGAGGGTTTATTGGAGCAGCCTGCCACGACGGTT

      G  A  E  D  K  N  W  H  Y  V  A  Q  G  D  R  S  L  L  P  L
7261  GGTGCCGAGGACAAGAATTGGCACTATGTCGCTCAGGGCGATCGTTCGCTGTTGCCGCTC

      E  V  F  D  D  G  F  T  T  V  F  H  F  P  G  N  V  R  I  P
7321  GAAGTCTTCGATGATGGATTTACGACGGTATTTCACTTCCCAGGTAATGTACGCATACCC

      S  I  Y  T  I  N  P  D  G  K  E  A  V  A  N  Y  S  V  K  G
7381  TCCATCTACACGATAAATCCGGATGGAAAGGAAGCTGTCGCTAACTATTCAGTCAAAGGG

      S  Y  V  E  I  S  S  V  S  R  G  W  R  L  R  D  G  H  T  V
7441  AGCTATGTCGAGATTTCTTCGGTTTCCCGTGGTTGGCGTCTGAGGGATGGCCACACGGTA
        ORF10a
      M  H  L  E  Y  R  L  R  S  R  R  P  Q  A  G  D  G  H  C  E
      L  C  I  W  N  T  A  Y  D  P  V  G  R  R  P  E  T  G  T  V
7501  TTATGCATTTGGAATACCGCCTACGATCCCGTCGGCCGCAGGCCGGAGACGGGCACTGTG
                                                ORF10b
      A  R  C  E  A  R  P  K  G  G  E  R  M  N  D  D  N  Q  Q  S
      R  P  D  V  K  R  V  L  K  E  V  R  G  *
7561  AGGCCCGATGTGAAGCGCGTCCTAAAGGAGGTGAGAGGATGAACGACGACAATCAGCAAT

      A  H  D  V  D  A  S  G  S  L  V  S  D  T  H  H  R  R  L  S
7621  CGGCGCATGATGTCGATGCGTCGGGGTCCCTGGTCTCCGACACACATCACCGGCGCCTTT

      G  A  Q  K  L  I  V  G  G  V  V  L  A  L  S  L  S  L  I  W
7681  CGGGGGCTCAAAAGTTGATCGTAGGAGGTGTAGTTCTCGCGCTATCACTTAGCCTCATTT

      L  G  G  R  E  K  K  E  N  G  D  A  P  P  S  T  M  I  A  T
7741  GGCTTGGCGGGCGTGAAAAGAAGGAAAACGGGGACGCACCCCCGTCAACCATGATCGCCA

      N  T  K  P  F  H  P  A  P  I  D  V  T  L  D  P  P  A  A  Q
7801  CGAACACCAAGCCATTTCATCCGGCTCCGATTGACGTTACACTTGATCCTCCGGCTGCCC

      E  A  V  Q  P  T  A  P  P  P  A  R  S  E  P  E  R  H  E  P
7861  AGGAAGCTGTTCAGCCGACTGCTCCTCCGCCAGCACGAAGTGAGCCGGAACGGCATGAGC

      R  P  E  E  T  P  I  F  A  Y  T  S  G  D  Q  G  T  S  K  R
7921  CGCGGCCGGAAGAAACACCGATTTTTGCGTACACCAGTGGTGATCAAGGGACCAGCAAGC

      V  Q  Q  G  E  T  D  R  R  R  E  G  N  G  E  D  S  P  L  P
7981  GCGTTCAACAAGGCGAAACGGACCGAAGACGCGAAGGCAATGGGGAAGACAGTCCTTTGC

      K  V  E  V  S  A  E  N  D  L  S  I  R  M  K  P  T  E  L  Q
8041  CGAAGGTCGAAGTGTCCGCCGAGAATGATCTCTCGATACGCATGAAGCCCACCGAGCTGC

      P  T  R  A  T  L  L  P  H  P  D  F  M  V  T  E  G  T  I  I
8101  AGCCCACCAGGGCTACGCTCTTGCCTCATCCCGACTTCATGGTGACGGAGGGGACGATTA

      P  C  I  L  Q  T  A  I  D  T  S  L  A  G  Y  V  K  C  V  L
8161  TTCCATGTATCTTGCAAACGGCAATCGACACCAGTCTGGCAGGCTATGTAAAATGCGTGT

      P  W  D  V  R  G  T  T  N  N  V  V  L  L  D  R  G  T  T  V
8221  TACCCTGGGATGTTCGTGGAACAACGAACAACGTTGTGCTTCTTGATCGCGGCACCACCG

      V  G  E  I  Q  R  G  L  Q  Q  G  D  A  R  V  F  V  L  W  D
8281  TTGTTGGCGAGATCCAGCGCGGTTTGCAACAGGGAGATGCGCGTGTTTTTGTGCTCTGGG

      R  A  E  T  P  D  H  A  M  I  S  L  A  S  P  S  A  D  E  L
8341  ATCGGGCGGAGACACCCGACCATGCCATGATTTCGCTTGCGTCACCAAGCGCTGACGAAC

      G  R  S  G  L  P  G  T  V  D  N  H  F  W  Q  R  F  S  G  A
8401  TCGGTCGCTCGGGATTGCCGGGCACCGTCGACAATCACTTCTGGCAGCGCTTTAGCGGGG

      M  L  L  S  V  V  Q  G  A  F  Q  A  A  S  T  Y  A  G  S  S
8461  CCATGCTCTTGAGTGTCGTCCAAGGTGCCTTCCAGGCAGCGAGCACCTACGCTGGCAGCT
```

Fig. 2

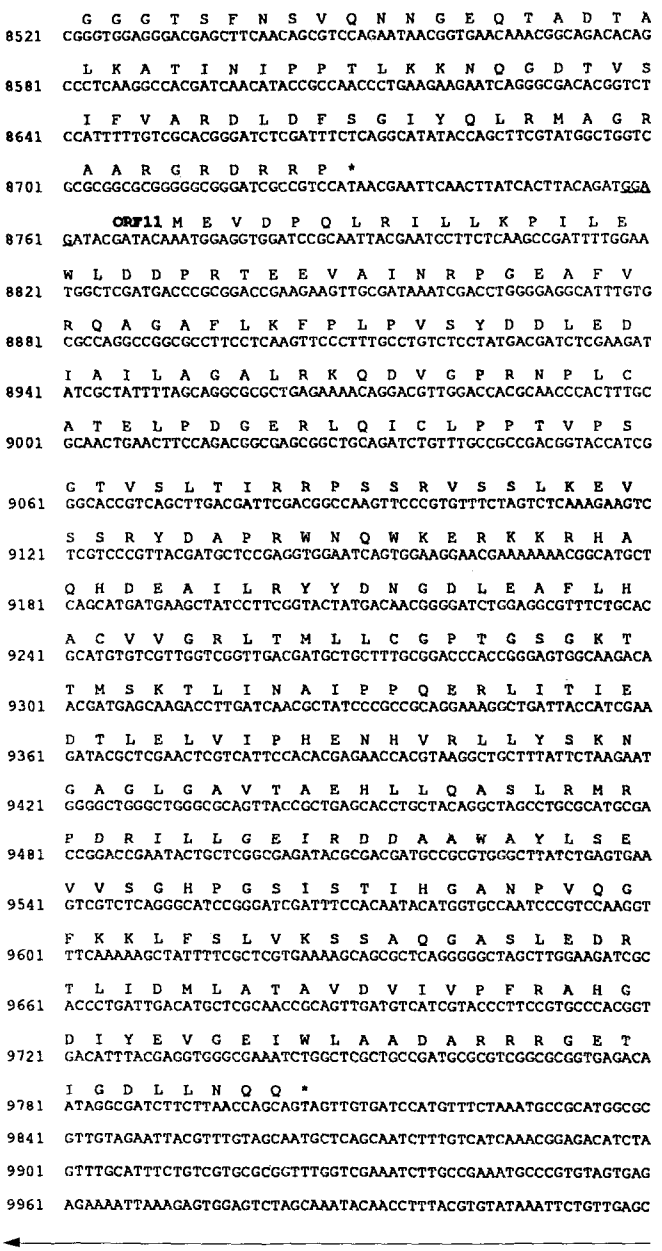

Fig. 2. Nucleotide sequence and predicted amino acid sequence of the pTiC58 *virB* operon. The amino acid sequence is shown in the one letter code. ATG methionine starts are indicated by *boldface* numbered ORFs. The two possible translation start sites for VirB6 and VirB10 are indicated a and b. Shine-Dalgarno ribosomal binding sequences are *underlined*. The sequence starts from the *virB* promoter region and ends with the *virB, virG* intercistronic region. The region from 9053 to 10020 is from Powell et al. (1987)

site may explain why this protein is produced at very low levels under its native signals in *E. coli*: a faint band migrating at 37 kDa is observed when plasmids containing *virB6* are expressed in *E. coli* maxi cells (Fig. 3B, lanes 4, 6). Since the nucleotide sequence of VirB6, initiating at the 6b site, predicts an ORF encoding a polypeptide of 31 kDa we suggest that VirB6 migrates slightly more slowly than predicted on SDS-PAGE.

The VirB7 ORF has a good SD sequence and attendant spacing to the ATG; however, due to its smaller molecular weight we were unable to confirm its size from expression studies. The VirB8 ORF has a rather weak SD sequence upstream of its predicted start. However, since the predicted size of this ORF is confirmed exactly by expression studies (Fig. 3B, lanes 2, 5, 6), we feel that the assignment of the start of this ORF is correct. It is intriguing to speculate that there may be no translation of ORF7; instead the relatively long intercistronic region between VirB6 and VirB8 may somehow promote translational initiation of VirB8.

Verification of predicted protein sizes of virB *ORFs in* E. coli

The predicted sizes of the *virB* ORFs were confirmed by visualization on SDS-PAGE, after expression using maxi cells and/or a T7 expression vector. To verify the predicted size of the first three ORFs, an in-frame fusion to the first ORF was made in the expression vector T7-7. The resulting fusion protein contains the entire VirB1 coding sequence plus an additional 18 amino acids from the vector and upstream *virB* sequences. This construct, pGK219 (Fig. 3A), also contains coding sequences for VirB2, VirB3 and the first 84 amino acids of VirB4. Transcription of these proteins from the T7 gene *10* promoter, is dependent on T7 RNA polymerase and is inducible by IPTG since transcription of the T7 RNA polymerase is under the control of a *tac* promoter. After induction of host cells carrying pGK219, and labeling with [^{35}S]methionine/cysteine, the predicted fusion to VirB1 at 30 kDa (i.e. VirB1 ORF plus 18 amino acids), VirB2, VirB3 and the first part of VirB4 can all be seen by SDS-PAGE when compared with controls without induction (Fig. 3B, lanes 9, 10, 11). An exact assignment of the individual VirB2 and VirB3 ORFs to the bands seen by SDS-PAGE is not possible since these polypeptides, and the portion of VirB4 present in this construct are of such similar size (12.3, 11.6 and 9.6 kDa respectively). However, three predicted polypeptides in this molecular weight range are visible, namely protein bands at 14, 13.8 and 12.5 kDa.

Many of the remaining ORFs were verified in *E. coli* maxi cells. Expression of pGVO310 (Fig. 3A) in maxi cells revealed six new protein bands as compared with the vector control, pBR322 (Fig. 3B, lanes 1, 2). Further analysis of smaller segments of the *virB* operon in maxi cells allowed us to assign most of these protein bands to specific *virB* ORFs. Expression in maxi cells was dependent on *Agrobacterium* sequences upstream of the first ORF as a construct lacking these sequences gave no expression (data not shown). Additionally, the level of expression of *virB*-encoded proteins did not change with orientation of the insert in the vector (pUC119) suggesting that vector promoters were not responsible for its transcription. Thus, the *A. tumefaciens virB* promoter apparently is recognized by *E. coli* RNA polymerase, allowing us to identify the sizes of *virB* specific proteins in *E. coli* maxi cells. Transcription from *vir* promoters in *E. coli* maxi cells has been observed previously by De Vos and Zambryski (1989) in a study of *virD*- and *virC*-encoded proteins. Since pGVO310 contains the entire *virB* operon in addition to *virA, tzs*, and most of *virG* it was necessary to dissect *virB* further in order to identify conclusively the sizes of the *virB* gene products.

Expression of pGK215 (Fig. 3A), containing the *virB* promoter (Das et al. 1986) followed by VirB1, VirB2, VirB3 and the first 84 amino acids of VirB4 showed *virB* specific bands at 12 and 14 kDa corresponding to VirB2, VirB3 and/or the VirB4 partial polypeptide when compared with a pUC119 control (data not shown). A protein band for

Table 1. Criteria for selection of translation initiation sites for VirB ORFs

ORF	1	2	3	4	5	6a	6b	7	8	9	10a	10b	11
SD	+	+	+	+	±	−	−	+	±	±	−	+	+
Spacing	+	+	+	+	±	−	−	+	−	±	−	+	+
Coupling	−	+	+	+	−	−	−	−	−	+	−	+	−
Predicted size	26.2	12.3	11.6	87.7	23.3	38.4	31.9	5.8	26.0	32.2	44.3	40.6	38.1
Observed Size	28.0	~13[a]	~13[a]	80.0	23.8	38.0	38.0	?	26.0	28.0	51.0	51.0	38.0

ORF, opening reading frame; SD, Shine and Dalgarno sequence; spacing, +indicates that end of SD is 5 to 13 nucleotides from the ATG start of translation, − indicates a deviation from this parameter; coupling indicates downstream ATG starts overlapped by the immediately upsteam stop codon; predicted size (kDa) is from the nucleotide sequence and observed size is from expression studies shown in Fig. 3

[a] Since we have not determined the order of migration of VirB2 and VirB3 they are given the same approximate size

VirB1 was not observed (Fig. 3B, lane 3). This result is surprising given the strong SD sequence preceding the start of this ORF; however it is possible that VirB1 migrates in the same area of the gel as β-lactamase from the vector and is thus obscured. Plasmid pGK216 with additional *virB* sequences (Fig. 3A) shows the expected proteins for VirB4, VirB5, and VirB6 at 80, 23 and 37 kDa respectively (Fig. 3B, lane 4). The VirB7 ORF of 5.8 kDa is too small to be resolved and must be considered a putative ORF at this time. Addition of the remaining *virB* sequences to pGK216 created pGK217 (Fig. 3A) which showed several additional proteins when expressed in maxi cells. This plasmid gave new bands at 26, 38 and 51 kDa. The 26 kDa band corresponds to VirB8 (predicted size is 26.0 kDa) and the other two bands correspond to VirB10 and VirB11. To determine the identity of VirB10 and VirB11 as either the 38 or 51 kDa proteins observed, a translation stop was created in VirB11 by filling in an *Asp*718 restriction site with Klenow DNA polymerase. This modification should result in the early termination of VirB11 producing a protein of approximately 12 instead of 38 kDa. When this construct (pGK218, Fig. 3A) was expressed in maxi cells the strong protein band at 38 kDa was absent, confirming that the 38 kDa protein is in fact VirB11 and the 51 kDa band is VirB10 (Fig. 3B, lanes 5, 6). The other predicted protein of 32.2 kDa (B9) was not observed in *E. coli* maxi cells. Potentially, this protein is not expressed well in maxi cells or may comigrate with a more intense band. To confirm the size of the VirB9 ORF we constructed pYT910, carrying the VirB9 and VirB10 ORFs. Induction expression of pYT910 produced two novel proteins migrating at 28.5 and 51 kDa on SDS-PAGE (Fig. 3B, lanes 11, 12). Since we showed the 51 kDa protein corresponds to VirB10, the 28.5 kDa band must correspond to VirB9 and this agrees well with the predicted mobility of the VirB9 ORF. Thus, we have confirmed the molecular weight assignments of 10 of the 11 *virB* ORFs. The only unconfirmed ORF is the very small VirB7 (5.8 kDa) which may be present, but is not detected in the gel system used here.

Comparison of the octopine and nopaline Ti plasmid virB *operons*

Figure 4 shows a comparison of the predicted ORFs between the two previously published sequences from octopine type plasmids and the sequence from the nopaline type plasmid presented here. There are two areas with notable differences. In the area of *virB4* (nopaline), both of the octopine sequences predict two ORFs with a combined size of 86.0 kDa (pTi15955) and 89.4 kDa (pTiA6) while the nopaline sequence predicts only one ORF of 87 kDa. All the ORFs mentioned are in the same frame. Expression studies described above demonstrate the presence of an 80 kDa protein coded from this region in nopaline. Additionally, Engstromm et al. (1987) have observed an 80 kDa protein, mapping to this region of the *virB* complementation group of pTiA6, which fractionated with the bacterial envelope. Thus, the prediction of an approximately 80 kDa protein from this area of the *virB* operon is supported both by in vivo studies in an octopine strain and by expression studies of nopaline *virB* sequences (Fig. 3B, lanes 2, 4, 5, 6). The other area of difference between octopine and nopaline sequences is in the region of *virB9* and *virB10* of the nopaline sequence. One of the published octopine sequences predicts only one larger ORF of 72 kDa (designated *virB10*; Ward et al. 1988) and the other report (Thompson et al. 1988) predicts two ORFs of 32.2 and 40.6 kDa (designated *virB9* and *virB10*) similar in size to those specified in nopaline designated *virB9* and *virB10*. Expression studies confirm that two proteins of 51 and 27 kDa apparent molecular weight are encoded by this region (Fig. 3B, lanes 2, 5, 6, 11).

All the predicted ORFs described here show high amino acid homology (all greater than 70% identity) to ORFs predicted from both of the published octopine sequences: VirB1, VirB2 and VirB3 of nopaline are homologous to VirB1, VirB2 and VirB3 of pTi15955 and pTiA6 (Fig. 4); VirB4 of nopaline is homologous to VirB4 and VirB5 of the octopine sequences; VirB5 nopaline is homologous to VirB6 octopine; VirB6 nopaline is homologous to VirB7 octopine; VirB7 nopaline is homologous to VirB8 of pTiA6 and the beginning of VirB8 of pTi15955; VirB8 nopaline is homologous to VirB9 of pTiA6 and VirB8 of pTi15955; VirB9 nopaline is homologous to VirB9 of pTi15955 and to the first half of VirB10 from pTiA6; VirB10 nopaline is homologous to VirB10 of pTi15955 and to the second half VirB10 from pTiA6; the VirB11 proteins from all three plasmids are homologous. While most of these homologies are in the range of 70% to 90%, there are stretches of perfect homology which may have been conserved due to functional constraints. A more detailed study of these areas may give clues to the functions of the VirB proteins. Based on our detailed expression studies that confirm the assignments of the VirB proteins, we propose that the numbering of the VirB proteins be as described here for the nopaline *virB* region.

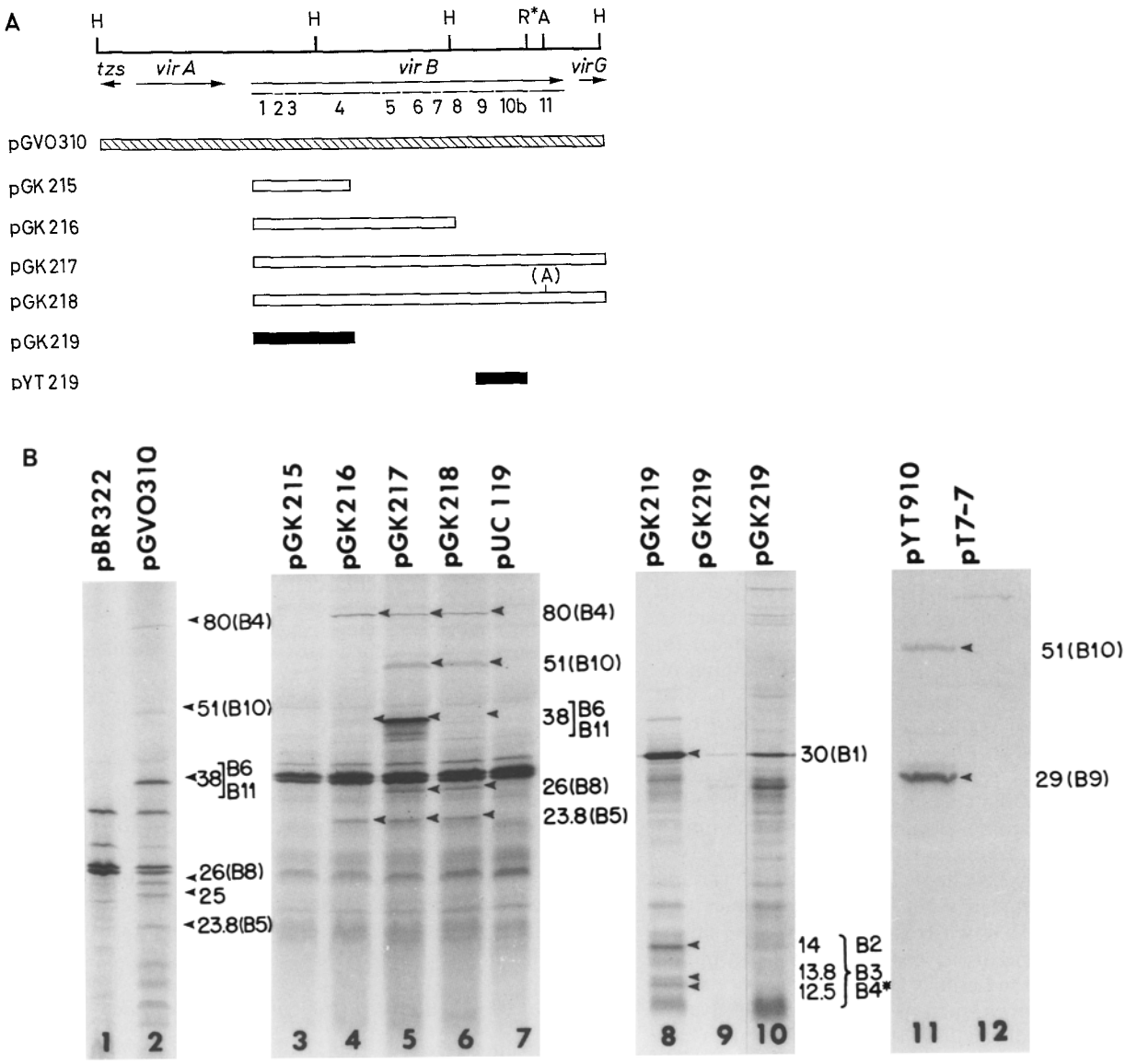

Fig. 3A and B. Expression of *virB* ORFs in *Escherichia coli*. **A** Restriction map of pGV0310 and positions of *virB, virG, virA* and *tzs* are shown. The *numbers* below the *virB* region indicate the ORFs predicted from the sequence in Fig. 2. *virB* sequences contained in subclones used for expression studies are shown as *boxes* beneath the restriction map; *diagonally lined* box represents pGV0310, *white boxes* indicate subclones used for expression in maxi cells, and *dark boxes* indicate subclones expressed under T7 RNA polymerase. **B** The plasmids used for expression are indicated above each lane. *Arrows* indicate the positions of VirB polypeptides and their sizes and ORF numbers are given to the side of each autoradiogram. The sizes of the polypeptides are given in kDa; these sizes were determined with reference to standard molecular weight markers (High Molecular Weight Markers from Bethesda Research Laboratories). Samples in lanes 1 and 2 were electrophoresed on a 13%–21% gradient gel, lanes 3 to 7 on a 12% gel, lanes 8 to 10 on a 15% gel, and lanes 11 and 12 on a 12% gel. Lanes 1 through 8 are radioactively labeled extracts of *E. coli* maxi cells. Lanes 8 to 12 are extracts of *E. coli* carrying plasmids which express VirB proteins under the control of a T7 promoter after induction by isopropyl-thio-β-galactoside (IPTG; lanes 8, 11, 12) and treatment with rifampicin. Lane 10 is a longer exposure of lane 9. The positions of VirB2, VirB3, and VirB4* (N-terminal region of VirB4 present in pGK219) in lane 10 are indicated together since their exact assignments to specific bands in this molecular weight region of the gel cannot be made

Hydropathy of nopaline VirB proteins

The work of Engstromm et al. (1987) has shown that at least three VirB proteins localize to the bacterial envelope suggesting that the VirB proteins may form a membrane channel or pore to export the T-strand from *Agrobacterium* cells. To determine whether the nopaline VirB proteins might be membrane associated, we examined the hydrophobicity of the VirB polypeptides. Hydropathy plot analysis, using the parameters of Engleman et al. (1986), revealed that 9 of the 11 ORFs have regions of hydrophobicity which could potentially permit interaction with a membrane. Only VirB4 and VirB11 do not have hydrophobic regions (of at least 10 amino-terminal residues or 20 internal amino acids). The hydrophilic nature of VirB4 is puzzling since Engstromm et al. (1987) have localized this protein to the bacterial envelope. The precise location of VirB4 and its characteristics need to be studied further. ORFs 1, 5, 7,

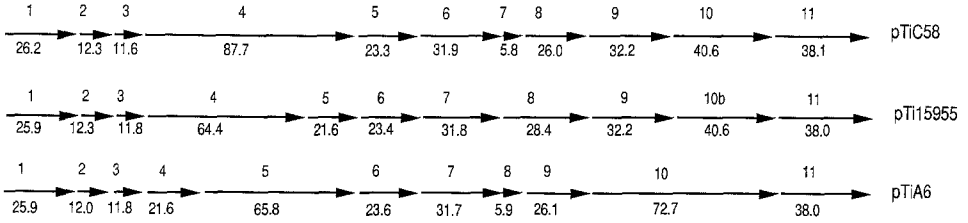

Fig. 4. Comparison of ORFs of *virB* operons from three different Ti plasmids. pTi15955 *virB* ORFs are from Thompson et al. (1988) and pTiA6 *virB* ORFs are from Ward et al. (1988). Numbers above each set of *arrows* indicate ORF designation and those below indicate predicted size (kDa)

and 9 have hydrophobic regions at their amino-termini suggesting possible export signal sequences. Additionally, ORFs 2, 3, 5, 6, 8 and 10 contain hydrophobic stretches of 20 or more internal amino acids which may be membrane spanning regions (data not shown). Overall VirB2 is the most hydrophobic: a large contiguous section of approximately 70 amino acids, representing 60% of the ORF, contains hydrophobic residues. VirB3 and VirB6 also show large regions of approximately 30 to 40 hydrophobic amino acids residues, while VirB8 and VirB10 each have one short area of hydrophobicity of approximately 20 amino acids. Note that ORF5 potentially contains a signal sequence for protein export as well as internal membrane spanning regions.

The amino-terminal regions of ORFs 1, 5 and 9 have features similar to those found in bacterial export signal sequences (Fig. 5). Such signal sequences are located at the amino-terminal end of the protein and consist of three distinct regions: (a) an amino-terminal portion containing 1 to 5 amino acids with an average net charge of +1.7 (von Heijne 1985); (b) a hydrophobic core region of variable length thought to form an alpha helix and interact with the membrane; and (c) a downstream region of 6 to 7 amino acids containing several polar residues. The hydrophobic and polar domains are sometimes separated by one or several helix breaking residues, such as proline or two glycines (von Heijne 1983). Additionally, these signal sequences have a consensus cleavage site at their end which is recognized by signal peptidase I. The consensus cleavage site is for alanine or glycine at the –1 position and for alanine, glycine or other small amino acids at the –3 position. This has been called von Heijne's –3, –1 rule (von Heijne 1983). Figure 5 shows signal sequences for VirB1, VirB5 and VirB9. The signal peptidase I cleavage sites are ser(–3), ala(–1) for VirB1, ala(–3), ala(–1) for VirB5 and ala(–3), ala(–1) for VirB9.

The amino-terminal end of ORF7 has features characteristic of a lipoprotein signal sequence (Fig. 5). Studies of lipoprotein signal sequences have shown some similarities to the signal peptidase I type described above. The main difference between the two signal sequences is in the signal peptidase recognition consensus sequence. In *E. coli*, lipoproteins are cleaved by signal peptidase II which recognizes a site similar to that recognized by other bacterial lipoprotein signal peptidases. The consensus cleavage site is for alanine or glycine at the –1 position and a glycerol modified cysteine at +1. ORF 7 contains a Gly, Cys (–1, +1) site shown in Fig. 5. It has been shown in *E. coli* that the amino acid at the +2 or +3 position in lipoproteins determines whether the lipoprotein is localized to the inner or outer membrane. A negatively charged amino acid at the +2

VirB ORF Signal Sequences

ORF 1 MLK+ATGPLSIILLASTCPSSGAAPLSFA▼EFNN

ORF 5 MKI+MQLVAAAMAVSLLSVGPARA▼QFVV

ORF 7 MKY+CLLCLALALGG▼CQTN

ORF 9 MTKK++AFLTLACLLFAAIGARA▼EDTP

Fig. 5. Potential protein export signal sequences in *virB* ORFs. Charged amino acids are indicated with a+. Hydrophobic core regions (described in text) are *underlined* and signal peptidase cleavage sites are shown with an *arrowhead*

or +3 position relative to the signal peptidase II cleavage site localizes lipoproteins to the inner membrane and uncharged amino acids localize the lipoproteins to the outer membrane (Yamaguchi et al. 1988). By these criteria the putative lipoprotein encoded by *virB7* would be localized to the outer membrane. Such an outer membrane localized lipoprotein may interact directly with the plant cell surface in the T-DNA transfer process.

Homology of VirB proteins to non-Vir polypeptides

A complete search of the NBRF, Genbank and EMBL databases has shown no significant homologies aside from those with previously reported *virB* genes (described above). VirB11 however is homologous to *Bacillus subtilis* ComG protein (David Dubnau, personal communication). *B. subtilis* cells endowed with functional competence genes (ComG) have the ability to take up single-stranded DNA from their environment. This may imply that some processes involved in single-stranded DNA uptake (*B. subtilis*) and export (*Agrobacterium*) are related.

Conclusions and prospects

We have presented the complete sequence of the *virB* operon of the nopaline plasmid pTiC58. The predicted sizes of the ORFs were verified for 10 of the 11 ORFs using detailed expression studies. A comparison of the nopaline *virB* operon with octopine *virB* operons shows that the *virB* operons are highly homologous. However, the assignment of ORFs between the three sequenced *virB* operons is not in agreement. This paper then clarifies the discrepancy among the sizes of VirB proteins. Hydropathy plot analysis has indicated 9 of the 11 VirB proteins are hydrophobic enough to interact with membranes further supporting hypotheses about the role of VirB as a membrane channel

or pore for the export of T-strands. The intergenic regions between many of the *virB* ORFs are small and suggestive of translational coupling; this may be a mechanism for efficient translation of the long *virB* operon by maximizing the time ribosomes are bound to the mRNA. The *virB* operon is a rich source of material for the study of this mode of protein synthesis. As the precise function of the *virB* gene products is not yet known, further experiments to determine the exact location and the characteristics of the VirB proteins are required.

Acknowledgements. We gratefully acknowledge Amie Franklin and Christian Tobias for deletion mapping, Yvonne Thorstenson for construction of pYT910, Frank Murrillo for graphics and all the members of our laboratory for helpful discussion. This work was supported by a grant from the Department of Energy (DE-FG03-88ER13882) to P.Z. and by an N.S.F. predoctoral fellowship to G.K.

References

Chen EY, Seeburg PH (1985) Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4:165–170

Christie P, Ward JE, Winans SC, Nester EW (1988) The *Agrobacterium tumefaciens virE2* gene product is a single-stranded DNA-binding protein that associates with T-DNA. J Bacteriol 170:2659–2667

Citovsky V, De Vos G, Zambryski P (1988a) A novel single-stranded DNA binding protein encoded by the *virE* locus, is produced following activation of the *A. tumefaciens* T-DNA transfer process. Science 240:501–504

Citovsky V, Howard E, Winsor B, Zambryski P (1988b) Proteins that mediate DNA transfer by *Agrobacterium tumefaciens* to plant cells: summary and perspectives. In: Staskawicz B, Alhquist P, Yoder O (eds) UCLA Symp Mol Cell Biol. Alan R. Liss, New York, pp 3–18

Das A (1988) The *A. tumefaciens virE* operon encodes a single-stranded DNA binding protein. Proc Natl Acad Sci USA 85:2909–2913

Das A, Yanofsky C (1984) A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. Nucleic Acids Res 12:4757–4768

Das A, Stachel SE, Ebert P, Allenza P, Montoya A, Nester E (1986) Promoters of *A. tumefaciens* Ti plasmid virulence genes. Nucleic Acids Res 14:1355–1364

Depicker A, DeWilde M, De Vos G, De Vos R, Van Montagu M, Shell J (1980) Molecular cloning of overlapping segments of the nopaline Ti plasmid of pTiC58 as a means to restriction endonuclease mapping. Plasmid 31:193–211

De Vos G, Zambryski P (1989) Expression of *Agrobacterium* nopaline-specific VirD1, VirD2, and VirC1 proteins and their requirement for T-strand production in *E. coli.* Mol Plant-Microbe Interact 2:43–52

Engleman DM, Steitz TA, Goldman A (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Chem 15:321–353

Engler G, Depicker A, Maenhaut R, Villarroel R, Van Montagu M, Shell J (1981) Physical mapping of DNA base sequence homologies between octopine and a nopaline Ti plasmid of *Agrobacterium tumefaciens*. J Mol Biol 152:183–208

Engstromm P, Zambryski P, Van Montagu M, Stachel SE (1987) Characterization of *Agrobacterium tumefaciens* virulence proteins induced by the plant factor acetosyringone. J Mol Biol 197:635–645

Gietl C, Koukolikova-Nicola Z, Hohn B (1987) Mobilization of T-DNA from *Agrobacterium* to plant cells involves a protein that binds single-stranded DNA. Proc Natl Acad Sci USA 84:9006–9010

Gold L (1988) Posttranscriptional regulatory mechanisms in *Escherichia coli.* Annu Rev Biochem 57:199–233

Herrella-Estrella A, Chen Z, Van Montagu M, Wang K (1988) VirD proteins of *Agrobacterium tumefaciens* are required for the formation of a covalent DNA/protein complex at the 5′ terminus of T-strand molecules. EMBO J 7:4055–4062

Howard EA, Winsor BA, De Vos G, Zambryski P (1989) Activation of the T-DNA transfer process in *Agrobacterium* results in the generation of a T-strand-protein complex: tight association of VirD2 with the 5′ ends of T-strands. Proc Natl Acad Sci USA 86:4017–4021

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Oppenheim D, Yanofsky C (1980) Translational coupling during expression of the trytophan operon in *Escherichia coli.* Genetics 95:785–795

Powell BS, Powell GK, Morris RO, Rogowsky PM, Kado CI (1987) Nucleotide sequence of the *virG* locus of the *Agrobacterium tumefaciens* plasmid pTiC58. Mol Microbiol 1:309–316

Ronson CW, Nixon BT, Ausubel FM (1987) Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell 5:579–581

Sancar A, Wharton RP, Selzer S, Kacinski BM, Clarke ND, Rupp WD (1981) Identification of the *uvrA* gene product. J Mol Biol 148:45–62

Shine J, Dalgarno L (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342–1346

Shinedling S, Swebilius B, Gayle M, Pribnow D, Jarvis E, Edgar B, Gold L (1987) Sequences and studies of bacteriophage T4rII mutants. J Mol Biol 195:471–480

Stachel SE, Nester EW (1986) The genetic and transcriptional organization of the *vir* region of the A6 Ti plasmid of *Agrobacterium tumefaciens*. EMBO J 5:1445–1454

Stachel SE, Zambryski PZ (1986) *virA* and *virG* control the plant-induced activation of the T-DNA transfer process of *Agrobacterium tumefaciens*. Cell 47:155–157

Stachel SE, Messens E, Van Montagu M, Zambryski PC (1985) Identification of the signal molecules produced by wounded plant cells which activate the T-DNA transfer process of *Agrobacterium tumefaciens*. Nature 318:624–629

Stachel SE, Timmerman B, Zambryski PZ (1986) Generation of single-stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens* to plant cells. Nature 322:706–712

Stachel SE, Timmerman B, Zambryski PZ (1987) Activation of *Agrobacterium tumefaciens vir* gene expression generates multiple single-stranded T-strand molecules from the pTiA6 T-region: requirement for 5′ virD products. EMBO J 6:857–863

Studier FW, Moffat BA (1986) Use of bacteriophage T7 RNA polymerase to direct selection of high level expression of cloned genes. J Mol Biol 189:113–130

Thompson DV, Melchers LS, Idler KB, Schilperroort RA, Hooykaas PJJ (1988) Analysis of the complete nucleotide sequence of the *Agrobacterium tumefaciens virB* operon. Nucleic Acids Res 16:4621–4636

Veluthambi K, Jayaswal RK, Gelvin SB (1987) Virulence genes A, G, and D mediate the double-stranded border cleavage of T-DNA from the *Agrobacterium* Ti plasmid. Proc Natl Acad Sci USA 84:1881–1885

Vieira J, Messing J (1987) Production of single stranded plasmid DNA. Methods Enzymol 153:3–11

von Heijne G (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133:17–21

von Heijne G (1985) Signal sequences the limits of variation. J Mol Biol 184:99–105

Ward E, Barnes WN (1988) VirD2 protein of *Agrobacterium tume-*

faciens is very tightly linked to the 5′ end of T-strand DNA. Science 242:927–930

Ward JE, Akiyoshi DE, Regiers D, Datta A, Gordon MP, Nester EW (1988) Characterization of the *virB* operon from an *Agrobacterium tumefaciens* Ti plasmid. J Biol Chem 263:5804–5814

Yamaguchi K, Fujio Y, Masayori I (1988) A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. Cell 53:423–432

Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR (1985) Mitochondrial origins. Proc Natl Acad Sci USA 82:4443–4447

Yanofsky MF, Porter SG, Young C, Albright LM, Gordon MP, Nester EW (1986) The *virD* operon of *Agrobacterium tumefaciens* encodes a site specific endonuclease. Cell 47:471–477

Zambryski PC (1988a) *Agrobacterium*-plant cell DNA transfer. In: Berg D, Howe M (eds) Mobile DNA. Am Soc Microbiol, Washington, DC, pp 309–333

Zambryski PC (1988b) Basic processes and underlying *Agrobacterium*-mediated DNA transfer to plant cells. Annu Rev Genet 22:1–30

Communicated by J. Schell

Received November 15, 1989