
Analysis of the complete nucleotide sequence of the *Agrobacterium tumefaciens* *virB* operon

David V. Thompson, Leo S. Melchers^{1*}, Ken B. Idler², Rob A. Schilperoort¹ and Paul J.J. Hooykaas¹

Agrigenetics Corporation, Advanced Research Division, 5649 East Buckeye Road, Madison, WI 53716, USA, ¹Biochemistry Laboratory, Department of Plant Molecular Biology, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands and ²Abbott Laboratories, Abbott Park, IL 60064, USA

Received March 1, 1988; Revised and Accepted April 21, 1988

Accession no. X06826

ABSTRACT

The complete nucleotide sequence of the *virB* locus, from the octopine Ti plasmid of *Agrobacterium tumefaciens* strain 15955, has been determined. In the large *virB*-operon (9600 nucleotides) we have identified eleven open reading frames, designated *virB1* to *virB11*. From DNA sequence analysis it is proposed that nearly all VirB products, i.e. VirB1 to VirB9, are secreted or membrane associated proteins. Interestingly, both a membrane protein (VirB4) and a potential cytoplasmic protein (VirB11) contain the consensus amino acid sequence of ATP-binding proteins. In view of the conjugative T-DNA transfer model, the VirB proteins are suggested to act at the bacterial surface and there play an important role in directing T-DNA transfer to plant cells.

INTRODUCTION

The pathogenic bacterium *Agrobacterium tumefaciens* genetically transforms plant cells by introducing a defined segment of DNA (T-region) from the tumor-inducing (Ti) plasmid into the plant genome (for recent reviews see 1, 2). Crown gall tumorigenesis results from the expression of T-DNA genes which encode enzymes for the production of the plant growth regulators auxin and cytokinin (3-5). Other T-DNA genes determine the production of certain specific compounds called opines in tumor cells (6, 7). The T-region does not encode functions for its transfer from the bacterium to the plant cell. In the Ti-plasmid the T-region is flanked by nearly identical 24 bp direct repeats, which form the cis-acting signals necessary for transfer (8-10). T-region transfer is mediated by products determined by virulence loci located elsewhere on the Ti-plasmid and on the *Agrobacterium* chromosome. The chromosomal virulence loci (*chvA*, *chvB*, *att* and *pscA* or *exoC*) specify the attachment of *Agrobacterium* to plant cells (11-14). The octopine Ti plasmid virulence (*vir*) region contains at least seven operons encoding trans-acting products (15-19) which are required for plant cell recognition and T-DNA transfer. The Vir-products which are absolutely essential are encoded by the *virA*, *virB*, *virD* and *virG* operons,

while the products determined by virC, virE and virF are only necessary for tumor induction on certain plant species.

Plant phenolic compounds such as acetosyringone and α -hydroxyacetosyringone specifically activate expression of the Ti plasmid vir-loci (20, 21) and trigger the T-DNA transfer process. Induction of vir-gene expression is regulated by proteins encoded by the virA and the virG locus (22,23). The VirA protein is an inner-membrane protein which most likely functions as a sensory protein for plant-signal molecules (24,25). The second regulatory component VirG is proposed to act as a positive regulatory protein which activates vir-gene expression (23,26). The two remaining vir-loci essential for tumor induction are virB and virD. A recent study of the virD locus shows that at least two proteins (VirD1 and VirD2) of the virD operon are involved in T-DNA processing (27). Together, these VirD1 and VirD2 proteins can induce a nick at a specific site within the T-region border repeats, which is followed by the generation of a single stranded T-DNA molecule (T-strand) in Agrobacterium. T-strand molecules are thought to be the T-DNA intermediates that are transferred to the plant cells during tumor induction (27, 28). The other locus essential for tumor induction is virB and comprises the largest vir-operon. However, to date no specific functions have been assigned to the virB locus. Recently, it was reported that three proteins encoded within the 5'-half of the virB locus are located in the cell envelope of acetosyringone induced Agrobacterium cells (29). Interestingly, the envelope localization of these VirB proteins suggests that they might be involved in the transfer of T-DNA across the Agrobacterium membrane to the plant cells.

In this report, we studied the nucleotide sequence of the entire virB-operon of the octopine type plasmid pTi15955. The virB operon spans 9.6 kb as defined by transposon mutagenesis and contains 11 open reading frames (ORFs). Some of the VirB proteins, as deduced from the DNA sequence, are extremely hydrophobic. Two VirB proteins, namely VirB4 and VirB11 contain the sequence characteristics of mononucleotide-binding-proteins. These findings are in line with a possible structural role of the virB encoded protein products in the T-DNA transfer process.

MATERIALS AND METHODS

Materials

Restriction endonucleases were purchased from either Promega Biotec or New England Biolabs and used according to suppliers recommendations. T₄

polynucleotide kinase was purchased from Pharmacia P.L. Biochemicals. (γ ³²P)ATP was purchased from New England Nuclear.

Strains and Plasmid Constructs

Agrobacterium tumefaciens strain 15955 (LBA 8255) was grown at 29°C in minimal medium (30) or LC-medium. Escherichia coli strain JM101, used for propagation of plasmid constructs, was grown in LC-medium. Plasmid isolation from Agrobacterium tumefaciens was done according to Koekman et al. (31), and from E.coli by the method of Birnboim and Doly (32). Standard recombinant DNA procedures were according to Maniatis et al. (33). A number of subclones were used to sequence across the virB region (See Fig. 1). Restriction fragments from pTi15955 were isolated from agarose gels by the method of Vogelstein et al. (34), using the "Gene clean" kit from Bio101. Vectors pUC19 or pIC19R were used for cloning (35, 36). The constructed vir-clones contain the following pTi15955 restriction fragments: 4.45 kb KpnI-BamHI fragment, pRAL3221; BamHI-14, pRAL3224; HindIII-34b+3, pRAL3229; BamHI-24, pRAL3232; BamHI-27, pRAL3240; SalI-12, pRAL3243 and SalI-13b, pRAL3244 (see Fig.1).

Nucleotide Sequencing

DNA sequence reactions were conducted according to the method of Maxam and Gilbert (37), as modified by Barker et al. (10). The DNA of the virB locus was sequenced on both strands over its entire length. Nucleic acid and amino acid sequences were analysed using the University of Wisconsin Genetics Computing Group programs.

RESULTS

Nucleotide sequence analysis of the virB locus

Extensive transposon mutagenesis of the octopine Ti plasmid revealed that the Vir-region contains seven transcriptional units (see Fig. 1) (15-19). Mutations in the VirB region, which spans about 9.6 kb, complement as a single locus indicating that virB consists of a large polycistronic operon. Fusions with a promoterless lac-operon demonstrated that expression of virB is inducible by specific plant phenolic compounds and that transcription of virB is clockwise towards the T-region (19,21, Melchers unpublished).

The nucleotide sequence of the entire virB operon is presented in Fig.2. There are eleven open reading frames, named virB1 to virB11, which fall within the VirB-region defined above. There are two possibilities for the start of the VirB10 coding region. Open reading frames begin at

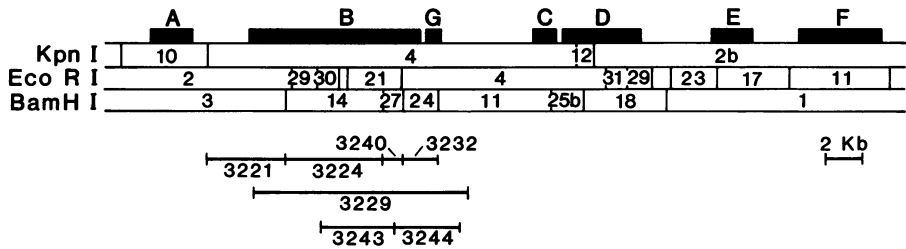


Figure 1. A physical map of the octopine plasmid pTi15955 Virulence region. Map positions of the seven different *vir*-loci are shown. The clones used for sequencing are shown below the restriction map.

nucleotide 7298 (*vir*B10a) and 7394 (*vir*B10b) and extend to nucleotide 8524 both in the same reading frame. The first start-codon (position 7298) overlaps with the coding region of *vir*B9 (overlap 32 amino acids), while the second start-codon (position 7394) overlaps with the stop-codon of ORF *vir*B9.

The nucleotide sequence of the *vir*B promoter region and transcription initiation site were reported previously (39). A comparison of the *vir*B promoter sequence of pTiA6 (39) to the promoter sequence of pTi15955 shows them to be identical. Analysis of the promoter region shows a -10 region (5'-GATAAT-3') with strong similarity to the *E.coli* consensus -10 sequence (5'-TATAAT-3'), while the -35 region of *vir*B (5'-TCGACT-3') contains only weak homology with the consensus -35 region (5'-TTGACA-3') of *E.coli* promoters (38). The *vir*B promoter region contains the hexanucleotide motifs (5'-GCAATT-3' and 5-CGAGTA-3'), identified by Das *et al.* (39). We identified upstream of the -35 region a nine base pair direct repeat (5'-CAATTGAAA-3') starting at nucleotide positions 36 and 56 of Fig.2, respectively. The palindromic hexanucleotide (CAATTG) was found also in the *vir*C/D promoter region (40) ; single base variants of this palindrome do also occur within all other inducible *vir*-promoters (our unpublished results).

Putative ribosome binding sites are found in front of nine of the eleven open reading frames. The sequence homology to the *E.coli* consensus ribosome binding sequence is shown in Table 1 (41). The nucleotide sequences preceding *vir*B5 and *vir*B8 do not show plausible matches to the ribosome binding site consensus of *E.coli*. Examination of the open reading frames shows that the translational start sites of *vir*B3, *vir*B4, *vir*B9 and *vir*B10b overlap the stopcodons of *vir*B2, *vir*B3, *vir*B8 and *vir*B9, respecti-

1 TTTTGTCCAGACGCCGCAACGGAAGTTTCCGTTCAATTGAAATCATAAAGAACAATTGAAATTTCGAACTMACCCGACCCCTCCGAGTAATCGTCAACATAAACAACAAGCCACTTCTTC

121 CAACCGGAGAGCCCGCGTTTACTGCGCCAGTAAAGAGATAAGCTTCTTTAGAGATCCGGGCGTCTTCCTCTCTGATCGTCTGCTCTGTTGCTGCAAGCGCTTCCGAGCCACTCTCTC

241 ATCTGCTGATTCAGACCATTTCTGCGCACTGCGCCACTTCCGCACTCACTTCCGGGGATAGCTAAGGTGAGACGTCGCTTGTATCTCTTACGCAATTGATGACACAGCCAC

361 GCGCGAAACGCTTTACTGCGGAAGATGACGACCGCAACGCAAGCGCAAGTCCGTCAGGCAAGGTCATGATGCAACGCGCATTCGCTGATGCGCTGATGCGAATAACTCTGCGAATTTTCGAT

481 GCTGCTTTACACGCTGCGCTGCGCCTGAAAGCGTCCGCTCAATCTCTGCGCTGCAAACTGTCGAAAGCTGTTATGACGCGCGCGAAACGATTGAGGAGACAGCAATTGCACTTCC

601 TCGGCGCATTCGCGTACAGCAACCGCGTAATTCATCGCGGCTTTTCCACACGCGCTACGTCGCAAAAGTTTAAACAAGCTGCTCAATCGCTGCTGCGCGCTTAACTGAGCTGCGCGCAAT

721 CGATGACAAAGCACTAAAGTCGGAGAATAGTCGGACGCTTTGGCGCTTATGACGCGCGCTCGCAAGAGATGGCGCTGCTCAATCGCTGCGCAACGCGCAAGCAAGCAAGCG

841 GAAATCCCGCAGCAGCAATTAAGTCTTATTCGACTTATGACTAAGGAGGTCCGCAATTGCGTTCGATGCTTTAGACAGATATGTTTAGACTTAAATCGCCTTCGCTTCCGATCGGATGATG

961 CCGCTGATATCGAGCCCGCCGCAAGTCCGCGCGCTGCGCTGCGCAATGGCAGCACTTCTCTGACGCGCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

1081 AATATATCGAATTCGCTTACTGCTGCTGCTGCGGAGTCACTGCGCGCTTTCGCGCTTCGCGCTATGCGCTATCGGGATCTCGCTGGAATTCGCGCGCGCTCGCTGCGCTGCGCTGCGCTGCG

1201 GCGCGCTGCTTACTGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTTTCGCGCT

1321 TTTCTGCGGTTCCGCTGACACTGCGCGCTTATTCATGATTTGCGCGCTTTCGCT

1441 TCACTGCTGAGCGGACTAGCAACCGCGCGCGCT

1561 TCGCGCGCGCGGAGGAGCACTGCT

1681 TGGATGCAATGAGCAATTCGCGCAATTCGCGGCTTCCGCTTTCGAGCCTTAGGAAATTTAAATCGCGCAATCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

1801 TGTCTGCAATATCTGCT

1921 CGAACTCTCTCGCAATGCGCTTTCGCT

2041 GATCAGGAACTGCGAAGATTTTGGCATTTTCGCT

2161 ACGCTGCAATGACTGCT

2281 AGACAGTACTGCT

2401 CTTGACTCGCTCGCAAGCGCAG

2521 CAAGCAATTCGCT

2641 TGTCCAGAAGTACTGCT

2761 AAATTCGCAAGCGCGCGCGCT

2881 ATTGCGCGCTGCTGCGAGCGCT

3001 ATTGCTGCTGCTGCGCCAGCGCAGCAGTATTTCGCACTTTCGCGCGCGCAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

3121 GATCTGCT

3241 GTTACTGCT

3361 CCGTGCATTCGCGCGCT

3481 ACCTCGCAATTCGCGCGCT

3601 CATGCT

3721 GAAATTCGCGCGCT

3841 CTTTCCGCGCGCT


```

7921  CGACCCGCATTTATCGCTAAGCGAAGCAATAATTCCTGCATCGCTGCAACCGCAATCGACACAATTTGCCAGCCTATCTGAAGCTCTCTCCCTCAGCATATTTCTGCAACAAC
    H  F  G  D  T  T  A  T  C  G  C  T  A  G  C  G  A  A  G  C  A  A  T  A  T  A  T  T  C  C  T  G  C  A  T  C  G  C  T  G  C  A  A  C  C  G  C  A  A  T  C  G  A  C  A  C  A  A  T  T  G  C  C  A  G  C  C  T  A  T  C  T  G  A  A  G  C  T  C  T  C  C  C  T  C  A  G  C  A  T  A  T  T  T  C  T  G  C  A  A  C  A  C
8041  GAACAATATCCTCTCTTAACTCCTGCGCAACCGCTTCCTGGCGAAATAGCGCTGCCTTCGCAACAGCGGAGATGAGCGCCCTTTTATGCTGTGGGATCGCCCGGACCCCTGACCATGC
    N  N  I  V  L  L  L  L  D  R  C  T  T  C  C  T  G  C  G  C  A  C  C  G  C  T  T  C  T  G  G  C  G  A  A  T  A  G  C  G  C  T  G  C  C  T  T  C  G  C  A  A  C  A  G  C  G  G  A  G  A  T  G  A  G  C  G  C  C  T  T  T  T  A  T  G  C  T  G  T  G  G  G  A  T  C  G  C  C  G  G  A  C  C  C  T  G  A  C  C  A  T  G  C  C  D  B  H  A
8161  GATGATCTGCTTAAACATCCCAAGCGCCGAGCACTCGCTGGCCAGGATTCGCCGCTCGGTTCACACACCACCTCTCGGAGCGCTTTAAGCGAGCCATCTCTTGAATCTCTTCAAGC
    H  I  S  L  L  I  S  L  F  S  A  D  E  F  V  S  I  F  V  A  R  D  L  D  F  F  G  V  V  Q  L  R  L  I  T  C  G  C  A  A  H  L  L  C  A  A  H  L  S  A  A  V  Q  G
8281  CCGCTTCGAGCCAGCTAGCAGCTGCGCAGCCTGGCGGATCGCTTCAACAAGCTTTCAAAAATAGCGTAAACAACAATGAGCAGCCCTTAGCGCCAGCCTCAACATAGC
    A  F  Q  A  A  A  S  T  A  C  S  S  G  C  E  C  H  S  F  N  S  T  Q  N  H  G  E  Q  A  T  E  A  L  A  L  R  A  A  E  T  F  F  A  T  N  I  F
8401  GCCAACCTTCAGAGAATATAGCGTACACCGCTTCCATTTCCTGGCCACCGACCTCGATTCTTTCCTTTACCGAGCTCCGCGTCACTCCGCGCCGCGCGCGCGGGGGAGAACCGCCG
    F  P  L  L  K  R  K  N  Q  G  D  T  V  S  I  F  V  A  R  D  L  D  F  F  G  V  V  Q  L  R  L  I  T  C  G  C  A  A  A  A  R  C  R  N  R  R
8521  CTCTTAATGATTTCAAAATTTCCGCTTAGACATAGCAGATCATTTCTAATGCAAGCTGATCGCAATACCGCTTTCTCTCAGCGCCTTTCGAAATGCGCTCATGACCGCCAGACATCAAG
    H  S  V  D  F  Q  L  R  R  E  L  L  D  D  B  A  T  C  T  A  E
    virB11
8641  AATTCCGATTAATCCGACTGAGAGCCATTTTCGCGCAACCGCGCATTTTACCAAGATGCGCTTACCGCTCTCTATGATGATTTGAGATATCCCTATTTTAGCGCCGCGCCGTCAG
    I  A  I  N  R  F  F  G  E  L  A  H  F  V  R  Q  A  G  A  C  I  F  T  C  R  H  R  L  P  V  S  V  T  D  D  L  S  E  D  I  A  I  L  S  L  A  G  C  A  T  L  R  R
8761  AAAGCAGATCTCGGACCACTTACCGCCCTCGCCCACTGAGCTCTGCTGTAAAGCTGACAAATCTCTCTCGGACCGGTTCCCTCCCGGACCGCTGACGCTTGCATTCGAGC
    K  Q  D  V  G  C  R  H  N  P  L  A  C  C  F  L  C  T  C  C  C  C  C  A  C  T  G  C  G  R  L  E  F  C  G  C  R  L  Q  L  R  L  I  T  C  G  C  A  A  I  H  S  R  Q  D  D  E  A  I  L  C  A  A
8881  GCCAGCTCCCGCTTTCTGCTTTAAAGAGTCTCCGCCCTTATGATCTCTCGAGCTGCACCGCTCCGACGACGACGAACGCAACCGCCAAAATCAGCATGTAAGCTATCTCTCAGCA
    F  S  T  R  V  S  G  L  R  E  V  S  R  Y  D  A  S  R  D  A  S  R  D  N  Q  Q  T  R  R  Q  V  I  D  A  S  C  A  A  C  A  G  C  C  C  A  A  A  T  C  A  G  C  A  T  G  A  A  G  C  T  A  T  C  T  T  C  A  G  C  A
9001  TTTTCAACACCGGATTTGGAAGCGTTCTGACGCAATGCTGTGCGCCACTGACGATCTGTGATCTGCGCCCTACCGGAAAGCGGACAGCAACATGAGCAAGACCTTGATCAGCCG
    I  D  N  C  D  L  E  L  H  A  C  E  F  V  R  Q  A  G  A  C  T  G  C  R  L  H  L  L  C  G  F  P  C  S  C  A  A  I  H  S  R  Q  D  D  E  A  I  L  C  A  A
9121  CATCCGAGCCGAGAAAGGCTAATTCACCAAGAGATACCTCGAACTCTGATTCACATGATATCAATGTTAGACTACTCTACTCAAAGAAGCGTCTCGGCTGGTCTCGGTGAGC
    I  P  P  Q  E  R  L  I  T  I  E  D  A  T  L  E  L  V  I  P  H  D  N  H  V  R  L  L  V  T  S  K  N  C  A  C  L  C  A  G  L  C  A  V  S  A
9241  CGAGCACTCTCCGACCAAGCAGTCTAGCTATGCGAGCGGACCGGATATGCTGGCGGAGATGCGGCAAGATCGACGATCGGCTATTGACTGAGTGAAGTGGCTCTCCGCGACATCGCGCAT
    H  E  L  L  Q  F  L  R  E  V  S  R  Y  D  A  S  R  D  D  A  A  C  H  R  D  D  A  A  C  A  G  C  T  G  C  T  T  C  G  A  G  T  G  A  G  T  G  G  C  T  C  C  G  G  A  C  A  T  C  G  G  C  A  T  G  C  A  T
9361  TTCACCAATCACGCGCGGAAATCCGATCCAAAGGATTCAAAGTACCTGTTTTCCCTGTGAAAGTACCGTCCAAAGCTGAGCTTGGAAAGATCGGACCACTCAATTGACATGCTCTCAACGGC
    S  A  T  I  H  C  C  G  A  N  F  C  C  A  T  C  C  A  G  G  A  T  C  A  A  E  A  C  T  G  T  T  T  C  C  T  G  T  G  A  A  A  G  T  A  G  C  G  T  C  C  A  A  G  C  T  G  A  G  C  T  T  G  G  A  A  G  A  T  C  G  G  A  C  C  A  C  T  A  T  T  G  A  C  A  T  G  C  T  C  T  C  A  C  G  C
9481  GATCGATTTTATTATTCGCTTCGCTCCATAGGACGTTTATGAGTACGGGACATCTCCCTCCGCGCGGACCGGACGACCGCGGCGGAGACCATAGGCAATCTCTTAATCAATAGT
    I  D  V  I  I  F  F  R  A  C  T  V  D  V  I  V  C  E  I  W  L  A  A  D  A  A  R  R  R  K  C  G  E  F  I  T  C  G  D  L  N  Q
9601  GCTGTAACTCGAAAGCGCTTCACTTGTAAACAGCATGAGAAATTTTGTGATAAAATTCAAATACTGGTTCGCATTTTGTTCATCGCCGCGCTCAGCGCGCAATCTCAGCAACTCGCCCAT
    H  V  V  H
    (virG)
9721  TAGCTGGAGTGTATTGACAT 9741
    (virG)

```

Figure 2. The nucleotide sequence of the virB operon. The complete DNA sequence of 9741 nucleotides derived from the clones shown in Fig. 1 is presented. The predicted amino acid sequences of the eleven open reading frames are shown below the DNA sequence in single letter code. The transcription initiation sites (39) at bp 101 and 103 are indicated with a star. The -10 and -35 region sequences are boxed. The arrows indicate the presence of a nine base pair direct repeat.

vely. This suggests that the expression of the subsets of ORFs virB2, virB3 and virB4 as well as those of virB8, virB9 and virB10b are translationally coupled (42). In the junction regions separating virB2-virB3 (UGAUG) and virB3-virB4 (UAAUG) the stop and start codons overlap just one base. Overlap of coding regions by one base exists also in the trp-operon of E.coli (43, 44) and in several gene pairs of bacteriophage lambda (45). A second type of overlap is present between the coding regions of virB8-virB9 and of virB9-virB10b (AUGA) whereby the stop and start codons overlap 2 bases. This phenomenon has also been observed in the genome of bacteriophage ϕ X174 (46) and in the virD-operon of Agrobacterium (47). The intercistronic regions in the virB operon are rather small, ranging in length from 0 (ORFs which abut one another) to 130 nucleotides (between virB7 and virB8), which is common in most polycistronic bacterial operons (48).

Table 1. Predicted Ribosome binding sites in virB.

B1	TAAGGAGaTA	-	4 bp	-ATG
B2	TAAGGAGGTc	-	7 bp	-ATG
B3	actGGcGGTa	-	4 bp	-ATG
B4	gAgGGAGagG	-	9 bp	-ATG
B5	attaccGGct	-	5 bp	-ATG
B6	TAAGGtaGga	-	4 bp	-ATG
B7	agttcAGGTc	-	6 bp	-ATG
B8	TttcccGcTG	-	1 bp	-ATG
B9	gtAGGccagG	-	7 bp	-ATG
B10a	gAgGGAtGgc	-	11 bp	-ATG
B10b	gAAGGgGGca	-	5 bp	-ATG
B11	atAGGAtaca	-	6 bp	-ATG
E.coli	TAAGGAGGTG	-	5-9 bp	-ATG

Nucleotides identical to the *E.coli* consensus (41) are capitalized.

Termination of virB transcription must occur within a region of 45 nucleotides (9599-9643) which is present between the last ORF (virB11) and the promoter region of the adjacent virG locus. At this 3'end of the virB operon there is no potential signal for factor-independent termination of virB transcription (49). From sequence analysis it turns out that the octopine T1 loci virB and virG are organized on the octopine T1 plasmid very close to each other. It has been observed that virG transcription is constitutive, but also inducible by plant-exudate to a higher level (19). If proper termination of virB-transcription occurs inefficiently, this will lead to higher levels of transcription of the adjacent virG operon upon induction of virB expression by plant signal molecules. This may in turn explain the inducibility of virG.

Proteins encoded by the virB operon

Computer analysis of the nucleotide sequence of virB revealed a coding capacity of eleven ORFs. The characteristics of the VirB proteins, as deduced from the nucleotide sequence i.e. number of amino acids, molecular weight and net charge are summarized in Table 2. Examination of the codon usage of the 11 virB-genes in addition to the ten already sequenced octopine T1 vir genes (virA, ref. 25; virG, ref. 26; virC1 and virC2, ref. 50, 51; virD1, virD2, virD3 and virD4, ref. 47,50; virE1 and virE2, ref. 52) shows that the *Agrobacterium* vir-genes utilize all codons with uniform frequency (data not shown). This is in contrast with the codon usage of *E.coli*, where certain codons are used rarely (for example, GGA (Gly) or CUA (Leu)) whereas others are used frequently (for example, GGU (Gly) or GUU (Val)) (53).

Table 2. Characteristics of the VirB proteins.

Vir protein	sequence location ORF	amino acids encoded	calculated MW	net charge
B1	164 - 880	239	25,952	-3
B2	898 - 1260	121	12,288	4
B3	1263 - 1586	108	11,759	2
B4	1589 - 3349	587	64,352	9
B5	3382 - 3954	191	21,633	-2
B6	3972 - 4631	220	23,450	-5
B7	4731 - 5615	295	31,771	-7
B8	5746 - 6516	257	28,362	1
B9	6516 - 7394	293	32,172	3
B10a	7298 - 8524	409	44,364	1
B10b	7394 - 8524	377	40,666	-3
B11	8567 - 9595	343	38,008	-7

During the tumor induction process, the T-DNA must cross the Agrobacterium membrane. Proteins localized in the bacterial inner membrane or outer membrane fraction are possible candidates which are functionally important in directing the T-DNA to the plant cell. In order to assign the possible cellular location of the proteins determined by the eleven virB ORFs we analyzed the distribution of hydrophobic and hydrophilic amino acid residues (see Fig. 3) using an algorithm developed by Kyte and Doolittle (54). Possible signal sequences were analyzed using the method of Von Heyne (55) to predict potential cleavage sites for signal peptidase. Interestingly, all VirB proteins except VirB3, VirB7, VirB10 and VirB11 contain at the N-terminus a putative signal peptide with a potential cleavage site as shown in Fig. 4. Features common to signal peptides precede the potential cleavage site in these VirB proteins, namely: a charged polar residue within the first 5 amino acids, a hydrophobic core sequence, and adjacent to the processing site a serine/alanine residue at position -3 while alanine is the most preferred residue at position -1. The proteins VirB3 and VirB7 lack a recognizable signal sequence although they are extremely hydrophobic (see Fig.3). Therefore, they are likely to be associated with the membrane of Agrobacterium as well.

A computer search using the Lipman and Pearson FASTP program (56) failed to reveal any sequence homology between the eleven VirB proteins (VirB1 to VirB11) and the proteins of the NBRF protein database (release 12, March 1987). Analysis of the VirB amino acid sequences in more detail identified a consensus sequence in VirB4 and VirB11 which is present in a

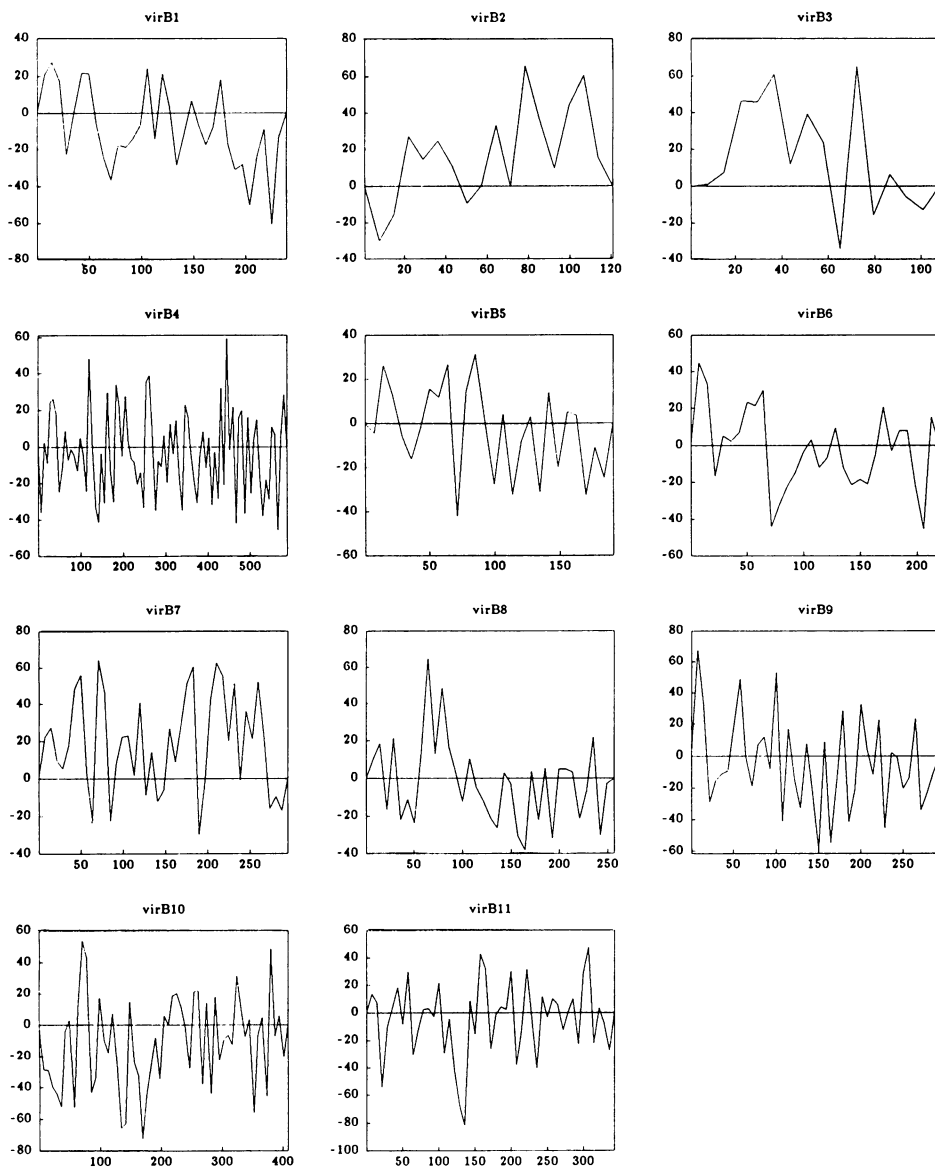


Figure 3. Hydrophobicity plots of the eleven VirB products (VirB1 to VirB11). The hydrophobicity profiles (values averaged over 7 amino acids) are plotted against the amino acid sequence positions by the method of Kyte and Doolittle (54). Values above the horizontal axis indicate hydrophobicity, while those below the axis indicate hydrophilicity.

Hydrophobic		cleavage site		S	protein
-20	-10	-1	+1		
MFKRSGSLSLALMSSFCSSSLA		/ TP		9.70	B1
MRCFERYRLHLNRLSLSNA		/ MM		4.77	B2
MLGASGTTTERS	GEIYLPYIGHLS	DHI VLLEDG SIMSIA / RI		6.56	B4
MTHLLEYEEVCAAAA		/ YL		4.39	B5
MKTQLIATVLTCSFLYIQPARA		/ QF		6.02	B6
MWGDGSLLRQIFSSAIRVDAMTGP	PEYAMLVARES LA / EH			6.51	B8
MTRKALFILACLFAAATGAEA		/ ED		10.69	B9

Figure 4. Putative signal sequences of VirB proteins. The signal peptide amino acid sequences were aligned from their potential cleavage site between residue -1 and residue +1. The scores (S-value) of the putative signal sequences were calculated using an algorithm of Von Heijne (55), and a window from -13 to +2. The predictive accuracy of this method is 75-80% .

wide variety of nucleotide-binding proteins (see Table 3). Crystallographic analysis of adenylate kinase and several other enzymes has shown that the conserved sequence (GXXXXGK) reflects a special strand motif that forms the phosphate binding region (57, 58). Many nucleotide binding proteins from both prokaryotes and eukaryotes retain this sequence, including kinases, ATP hydrolases, ATP-binding subunits of periplasmic transport systems (59) and the GTP-binding *ras* gene product p21. The proteins aligned in Table 3 all possess the consensus sequence of a nucleotide binding site although besides this region they lack significant homology with the proteins VirB4 and VirB11. It is important to note that most bacterial proteins that bind nucleotides, such as elongation and initiation factors, RecA and UvrD, also retain this short consensus sequence but share no additional homology.

DISCUSSION

The *virB* operon of *Agrobacterium tumefaciens* is essential for tumorigenesis. Homology studies of different types of Ti and Ri plasmids have shown that the *virB* locus is the most conserved part within the virulence regions of these plasmids (60, 61). The present nucleotide sequence analysis demonstrates that the octopine Ti *virB* operon contains eleven open reading frames. From the analysis of the VirB amino acid sequences, we suggest that most of the VirB proteins are membrane proteins. Signal sequences, predicted by an algorithm of Von Heijne (55), are identified in the N-terminus of the proteins, VirB1, VirB2, VirB4, VirB5,

Table 3. Alignment of the predicted amino acid sequence of VirB4 and VirB11 with various prokaryotic proteins comprising the consensus sequence which is characteristic of a mono-nucleotide binding site.

Protein	Species		Sequence
VirB4	A.tumefaciens	427	VGMTAIFGGPIGRGKTTLMM
VirB11	A.tumefaciens	162	RLTMLLCGPTGSGKTTMSK
HisP	S.typhimurium	32	GDVISIIGSSGSGKSTFLR
MalK	E.coli	29	GEFVVFVGGPSSGCGKSTLLR
PstB	E.coli	36	NQVTAFIGPSSGCGKSTLLR
NodI	R.leguminosarum	38	GECFLLGPNAGKSTITR
HlyB	E.coli	495	GEVIGIVGRSGSGKSTLTK
ATPase β	E.coli	143	GGKVGLFGGAGVGKTVNMM
ATPase α	E.coli	162	GQRELIIGDRQTGKTALAI
EF-Tu	E.coli	12	HVNVGTIGHVDHGKTTLTA
UvrD	E.coli	22	RSNLLVLGAGSGKTRVLV
RecA	E.coli	59	GRIVEIYGPESSGKTTLL

The consensus sequence (67) is boxed. See ref. 59 and 68 for references to these sequences and for more extensive listings. The number to left of each sequence is the position of the first amino acid shown within the complete protein.

VirB6, VirB8 and VirB9. In addition, the hydropathy profiles of VirB3 and VirB7 predict that these extremely hydrophobic proteins are associated with the Agrobacterium membrane, although they lack an obvious signal peptide. It has been shown that three VirB products of approximate molecular weights 33,000 (B33), 80,000 (B80) and 25,000 (B25) fractionate with the cell envelope of acetosyringone induced cells (29). From the relative location of their coding regions within the virB locus and the nucleotide sequence in this report we can conclude that B33, B80 and B25 correspond to VirB1 (MW 25,952), VirB4 (MW 64,352) and VirB6 (MW 23,450), respectively. The membrane location of VirB6 was recently confirmed. VirB6-PhoA hybrid proteins consisting of the first 207 amino acids of VirB6 fused to the carboxyl-terminal portion of alkaline phosphatase (PhoA) confer on Agrobacterium strong alkaline phosphatase activity (Melchers et al. unpublished). The reason for the discrepancy in the predicted and apparent molecular weights in the case of VirB1 and VirB4 is unclear, but has been observed in other proteins. For example, the second protein of the virD-operon is predicted to be 47.4 kDa but migrates in SDS-polyacrylamide gels with an apparent molecular weight of approximately 56 kDa (62).

Similar aberrant mobilities on gels have been observed for the products VirC1, VirE2 and several other proteins (51, 52, 63). Hence, both the amino acid sequence analysis of VirB1, VirB4 and VirB6, and the data on their cellular location clearly indicates that these VirB proteins are Agrobacterium membrane proteins.

After induction of vir-gene expression single-stranded T-DNA molecules, so called T-strands, are generated in Agrobacterium (27, 28). It is likely that the T-strand is the T-DNA intermediate molecule which A. tumefaciens mobilizes to the plant cell. It is interesting to speculate that T-DNA transfer is established by conjugation between A. tumefaciens and the plant cell, analogous to the conjugative transfer of plasmid DNA between prokaryotes. This predicts that several vir-encoded proteins are involved in this conjugative process, such as proteins that form pilus-like structures, contribute to conjugal DNA metabolism or regulation of the expression of the transfer operon (64). The filamentous F pili of E.coli are the best known example of conjugative pili which promote cell-to-cell contact during bacterial conjugation. F pilus formation is a complex process and requires at least 14 genes in the F transfer (tra) region (64), although the F pilus has an apparently simple structure (65). The large virB operon is a good candidate for a pilus operon in Agrobacterium, although there is no significant sequence homology between the VirB proteins and any of the known Tra-products (TraA, TraL, TraE, TraM) (66) or E.coli pili proteins (for example: PapA, PapG, PapH, FimF, FimG, FimH). The (membrane) proteins VirB2 (121 a.a.) and VirB3 (108 a.a.) correspond only in size to the TraA protein (119 a.a.), which following cleavage by signal peptidase forms the structural subunit of F pili.

It is interesting that a potential ATP-binding site (GXXGXGKT) is present in VirB4 (a.a. position 433) and VirB11 (a.a. position 169). The presence of an ATP-binding subunit is reported to be a common feature of cytoplasmic components from different periplasmic transport systems (for example: PstB, E.coli phosphate transport; HisP, S.typhimurium histidine transport; MalK, E.coli maltose transport). The identification of the ATP-binding consensus sequence in a number of other proteins, e.g. UvrD (DNA dependent ATPase), NodI (R.leguminosarum nodulation), RecA (ATP-dependent unwinding of double stranded DNA) and HlyB (haemolysin secretion), implies that ATP-hydrolysis is coupled to a variety of distinct biological processes (59). In our view, a possible function of the membrane protein VirB4 might be to provide the energy, via hydrolysis of ATP, for

translocation of virulence proteins or for the transfer of a T-DNA-protein complex across the Agrobacterium membrane. In view of the conjugative T-DNA transfer model it is interesting to speculate that VirB4 and leader peptidase are cooperatively involved in the transport of (virulence) proteins. Proteins essential for the assembly of pilus-like structures have to be exported. In addition, other proteins involved in the alteration of the bacterial cell surface are likely to play an essential role in the transfer of the T-DNA across the cell wall. Further characterization of the proteins VirB4 and VirB11 (e.g. photoaffinity labelling with ATP-analogues) will be required to confirm the identification of the ATP-binding sequence. To understand the functions of all the VirB proteins and their roles during the plant cell transformation process first the cellular location of all VirB proteins have to be established. In future research antibodies raised against each specific virB product will be used to identify their cellular location within acetosyringone induced Agrobacterium cells.

ACKNOWLEDGEMENTS

We thank Dr. Kees Rodenburg and Dr. Ron van Veen for critical reading of the manuscript. We are grateful to Adry van Es for typing the manuscript. This work was supported by the Agrigenetics Corporation and by the Netherlands Foundation of Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

* To whom correspondence should be addressed.

REFERENCES

1. Melchers, L.S. and Hooykaas, P.J.J. (1987) In: Oxford Surveys of Plant Molecular and Cell Biology 4, 167-220. Ed. Mifflin, B.J. Oxford University press.
2. Nester, E.W., Gordon, M.P., Amasino, R.M. and Yanofsky, M.F. (1984) Ann. Rev. Plant Physiol. 35, 387-413.
3. Schröder, G. Waffenschmidt, S., Weiler, E.W. and Schröder, J. (1984) Eur. J. Biochem. 138, 387-391.
4. Thomashow, L.S., Reeve, S. and Thomashow, M.F. (1984) Proc. Natl. Acad. Sci. USA 81, 5071-5075.
5. Akiyoshi, D.E., Klee, H., Amasino, R.M., Nester, E.W. and Gordon, M.P. (1984) Proc. Natl. Acad. Sci. USA 81, 5994-5998.
6. Bomhoff, G., Klapwijk, P.M., Kester, H.C.M., Schilperoort, R.A., Hernalsteens, J.P. and Schell, J. (1976) Molec. Gen. Genet. 145, 177-181.
7. Guyon, P., Chilton, M.-D., Petit, A. and Tempé, J. (1980) Proc. Natl. Acad. Sci. USA 77, 2693-2697.
8. Simpson, R.B., O'Hara, P.J., Kwok, W., Montoya, A.M., Lichtenstein, C.,

- Gordon, M.P. and Nester, E.W. (1982) *Cell* 29, 1005-1014.
9. Yadav, N.S., Van der Leyden, J., Bennett, D.R., Barnes, W.M. and Chilton, M.-D. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6322-6326.
 10. Barker, R.F., Idler, K.B., Thompson, D.V. and Kemp, J.D. (1983) *Plant Mol. Biol.* 2, 335-350.
 11. Douglas, C.J., Staneloni, R.J., Rubin, R.A. and Nester, E.W. (1985) *J. Bacteriol.* 161, 850-860.
 12. Matthijsse, A.G. (1987) *J. Bacteriol.* 169, 313-323.
 13. Thomashow, M.F., Karlinsey, J.E., Marks, J.R. and Hurlbert, R.E. (1987) *J. Bacteriol.* 169, 3209-3216.
 14. Cangelosi, G.A., Hung, L., Puvanesarajah, V., Stacey, G., Ozga, D.A., Leigh, J.A. and Nester, E.W. (1987) *J. Bacteriol.* 169, 2086-2091.
 15. Hille, J., Klasen, I. and Schilperoort, R.A. (1982) *Plasmid* 7, 107-116.
 16. Klee, H., White, F.F., Iyer, V.N., Gordon, M.P. and Nester, E.W. (1983) *J. Bacteriol.* 153, 878-883.
 17. Hille, J., Van Kan, J. and Schilperoort, R.A. (1984) *J. Bacteriol.* 158, 754-756.
 18. Hooykaas, P.J.J., Hofker, M., Den Dulk-Ras, H. and Schilperoort, R.A. (1984) *Plasmid* 11, 195-205.
 19. Stachel, S.E. and Nester, E.W. (1986) *EMBO J.* 5, 1445-1454.
 20. Okker, R.J.H., Spaik, H., Hille, J., Van Brussel, T.A.N., Lugtenberg, B. and Schilperoort, R.A. (1984) *Nature* 312, 564-566.
 21. Stachel, S.E., Messens, E., Van Montagu, M., Zambryski, P. (1985) *Nature* 318, 624-629.
 22. Stachel, S.E. and Zambryski, P.C. (1986) *Cell* 46, 325-333.
 23. Winans, S.C., Ebert, P.R., Stachel, S.E., Gordon, M.P. and Nester, E.W. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8278-8282.
 24. Leroux, B., Yanofsky, M.F., Winans, S.C., Ward, J.E., Ziegler, S.F. and Nester, E.W. (1987) *EMBO J.* 6, 849-856.
 25. Melchers, L.S., Thompson, D.V., Idler, K.B., Neuteboom, S.T.C., De Maagd, R.A., Schilperoort, R.A. and Hooykaas, P.J.J. (1987) *Plant Mol. Biol.* 9, 635-645.
 26. Melchers, L.S., Thompson, D.V., Idler, K.B., Schilperoort, R.A. and Hooykaas, P.J.J. (1986) *Nucleic Acids Res.* 14, 9933-9942.
 27. Stachel, S.E., Timmerman, B. and Zambryski, P. (1987) *EMBO J.* 6, 857-863.
 28. Van Haaren, M.J.J., Sedee, N.J.A., Schilperoort, R.A. and Hooykaas, P.J.J. (1987) *Nucleic Acids Res.* 15, 8983-8997.
 29. Engström, P., Zambryski, P., Van Montagu, M. and Stachel, S.E. (1987) *J. Mol. Biol.* 197, 635-645.
 30. Hooykaas, P.J.J., Roobol, C. and Schilperoort, R.A. (1979) *J. Gen. Microbiol.* 110, 693-701.
 31. Koekman, B.P., Hooykaas, P.J.J. and Schilperoort, R.A. (1980) *Plasmid* 4, 184-195.
 32. Birnboim, H.C. and Doly, J. (1979) *Nucleic Acids Res.* 7, 1513-1523.
 33. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
 34. Vogelstein, B. and Gillespie, D. (1979) *Proc. Natl. Acad. Sci. USA* 76, 615-619.
 35. Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* 26, 101-106.
 36. Marsh, J.L., Erfle, M. and Wykes, E.J. (1984) *Gene* 32, 481-485.
 37. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
 38. Hawley, D.K. and McClure, W.R. (1983) *Nucleic Acids Res.* 11, 2237-2255
 39. Das, A. Stachel, S., Ebert, P., Allenza, P., Montoya, A. and Nester, E. (1986) *Nucleic Acids Res.* 14, 1355-1364.
 40. Tate, M.E. (1987) *Nucleic Acids Res.* 15, 6739.

Nucleic Acids Research

41. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 77, 7117-7121.
42. Das, A. and Yanofsky, C. (1984) *Nucleic Acids Res.* 12, 4757-4768.
43. Nichols, B. and Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. USA* 76, 5244-5248.
44. Platt, T. and Yanofsky, C. (1975) *Proc. Natl. Acad. Sci. USA* 72, 2399-2403.
45. Sanger, F., Coulson, A., Hong, G., Hiu, D. and Petersen, G. (1982) *J. Mol. Biol.* 162, 729-773.
46. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M. and Smith, M. (1977) *Nature* 265, 678-695.
47. Porter, S.G., Yanofsky, M.F. and Nester, E.W. (1987) *Nucleic Acids Res.* 15, 7503-7517.
48. Kozak, M. (1983) *Microbiol. Rev.* 47, 1-45.
49. Brendel, V. and Trifonov, E.N. (1984) *Nucleic Acids Res.* 12, 4411-4427.
50. Thompson, D.V., Idler, K.B., Melchers, L.S., our unpublished results.
51. Yanofsky, M.F. and Nester, E.W. (1986) *J. Bacteriol.* 168, 244-250.
52. Winans, S.C., Allenza, P., Stachel, S.E., McBride, K.E. and Nester, E.W. (1987) *Nucleic Acids Res.* 15, 825-837.
53. Sharp, P.M. and Li, W.H. (1986) *Nucleic Acids Res.* 14, 7737-7749.
54. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.
55. Von Heyne, G. (1986) *Nucleic Acids Res.* 14, 4683-4690.
56. Lipman, D.J. and Pearson, W.R. (1985) *Science* 227, 1435-1441.
57. Pai, E.F., Sachsenheimer, W., Schirmer, R.H. and Schultz, G.E. (1977) *J. Mol. Biol.* 114, 37-45.
58. Fry, D.C., Kuby, S.A. and Mildvan, A.S. (1986) *Proc. Natl. Acad. Sci. USA* 83, 907-911.
59. Higgins, C.F., Hiles, I.D., Salmond, G.P.C., Gill, D.R., Downie, J.A., Evans, I.J., Holland, I.B., Gray, L., Buckel, S.D., Bell, A.W. and Hermodson, M.A. (1986) *Nature* 323, 448-450.
60. White, F.F. and Nester, E.W. (1980) *J. Bacteriol.* 144, 710-720.
61. Risuleo, G., Battistoni, P., Costantino, P. (1982) *Plasmid* 7, 45-51.
62. Yanofsky, M.F., Porter, S.G., Young, C. Albright, L.M. Gordon, M.P. and Nester, E.W. (1986) *Cell* 47, 471-477.
63. Merrick, M.J. and Gibbins, J.R. (1985) *Nucleic Acids Res.* 9, 309-314.
64. Willetts, N.S. and Skurray, R. (1980) *Ann. Rev. Genet.* 14, 41-76.
65. Folkhard, W., Leonard, K.R., Malmsey, S., Marvin, D.A., Dubochet, J., Engel, A., Achtman, M. and Helmuth, R. (1979) *J. Mol. Biol.* 130, 145-160.
66. Finlay, B.B., Frost, L.S. and Paranchych, W. (1986) *J. Bacteriol.* 168, 990-998.
67. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO J.* 1, 945-951.
68. Möller, W. and Amons, R. (1985) *FEBS* 186, 1-7.