JOURNAL OF BACTERIOLOGY, Feb. 1990, p. 531–537
0021-9193/90/020531-07$02.00/0

Vol. 172, No. 2

The Regulatory VirG Protein Specifically Binds to a *cis*-Acting Regulatory Sequence Involved in Transcriptional Activation of *Agrobacterium tumefaciens* Virulence Genes

SHOUGUANG JIN, THOMAS ROITSCH, PETER J. CHRISTIE,† AND EUGENE W. NESTER*

Department of Microbiology, SC-42, University of Washington, Seattle, Washington 98195

Received 7 August 1989/Accepted 1 November 1989

Virulence genes of *Agrobacterium tumefaciens* are induced in parallel in the presence of plant phenolic compounds such as acetosyringone and the two regulatory *vir* genes *virA* and *virG*. In this study we identified a *cis*-acting regulatory sequence in the 5′-noncoding region of the *virE* operon that is essential for this activation. To do this, we constructed a series of deletion mutants by using exonuclease *Bal* 31. Western blot (immunoblot) analysis showed that the 70 base pairs upstream of the transcriptional start site were sufficient for full *virE* gene induction. A conserved dodecadeoxynucleotide sequence (*vir* box), which was previously identified in the nontranscribed sequences of all *vir* genes, was located at 5′ end of the minimum required promoter sequence. Deletion of this *vir* box only completely abolished induction of the *virE* gene. This demonstrates that the *vir* box functions as an upstream regulatory sequence. To study the role of the VirG protein in the activation process, we overproduced the native-sized VirG protein in *Escherichia coli* by fusing the *lacZ′* start codon ATG with the second *virG* codon AAA using site-directed mutagenesis. The VirG protein was purified and renatured from *E. coli* and was shown to bind to a specific sequence in two *vir* gene promoters. Footprinting analysis of the *virE* and *virB* promoters identified the 12-base-pair *vir* box as the VirG-binding core sequence.

Agrobacterium tumefaciens is a phytopathogenic gram-negative soil bacterium that is able to genetically transform susceptible plant cells with the production of crown gall tumors. This system is proving to be useful for studying plant-microbe interactions and creating transgenic plants for practical applications. During transformation, a specific segment of the large tumor-inducing (Ti) plasmid, the T-DNA, is transferred to the plant cell nucleus (for reviews, see references 3 and 40). The T-DNA transfer process requires the expression of the Ti plasmid-encoded virulence (*vir*) genes consisting of at least seven genetically identified operons, *virA*, *-B*, *-G*, *-C*, *-D*, *-E*, and *-H* (*pinF*), in octopine strains (13, 29). Two of the *vir* genes, *virA* and *virG*, are required for the induction of all *vir* genes in response to plant exudates containing low-molecular-weight phenolic compounds such as acetosyringone. Based on protein sequence similarities, *virA* and *virG* gene products have been identified as members of a large number of two-component regulatory systems, including EnvZ-OmpR, NtrB-NtrC, and CheA-CheY/B (16, 35). These proteins link the expression of sets of genes with specific environmental stimuli. According to the present model (27), which has been confirmed for some of the homologous proteins, the first protein (sensor) is autophosphorylated in response to the environmental signal. The information is then transferred via phosphoryl group transfer to the second component (activator), which in turn activates a series of genes which respond to the environment. Specific gene expression is mediated by the binding of the activator to regulatory sequences 5′ of the promoter.

Conserved sequences in the 5′-noncoding region of all *vir* genes have been identified by DNA sequence analysis (9, 36). The consensus dodecadeoxynucleotide sequence, TNC AATTGAAAPy (where Py is a pyrimidine), was postulated to be a *cis*-acting regulatory sequence that serves as a VirG protein-binding site (36). To provide experimental data on this model and to identify other potential *cis*-acting regulatory sequences, we have carried out a functional analysis of the promoter region of the *virE* operon (6, 7, 34). We have also analyzed the interaction of the VirG regulatory protein with the upstream region of the *virE* and *virB* promoters that are essential for their activation.

MATERIALS AND METHODS

Enzymes and reagents. Restriction endonucleases and DNA-modifying enzymes were purchased from either New England BioLabs, Inc. (Beverly, Mass.), or Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), and were used according to the recommendations of the suppliers. The Sequenase DNA sequencing kit was obtained from U.S. Biochemical Corp. (Cleveland, Ohio), and a oligonucleotide-directed mutagenesis kit was obtained from Amersham Corp. (Arlington Heights, Ill.). All ^{32}P-labeled nucleotides were supplied by Dupont, NEN Research Products (Boston, Mass.). The Klenow fragment of DNA polymerase I, polynucleotide kinase, and T4 DNA ligase were obtained from Bethesda Research Laboratories; *p*-aminobenzyl 1-thio-β-D-galactopyranoside-agarose was supplied by Sigma Chemical Co. (St. Louis, Mo.); poly(dI-dC)(dI-dC) and DNase I were from Pharmacia Fine Chemicals (Piscataway, N.J.); and Immobilon membranes were from Millipore Corp. (Bedford, Mass.).

Strains and plasmids. *A. tumefaciens* Mx358 was a Tn*3*HoHo1 insertional *virE* mutant of strain A348 (29). Strain A136 was a derivative of strain C58 which was cured of its nopaline Ti plasmid. *Escherichia coli* TG1 was supplied with the oligonucleotide mutagenesis kit from Amersham; strain DH5α was purchased from Bethesda Research Laboratories. Phagemids pTZ18R and pTZ19R were purchased

* Corresponding author.
† Present address: Department of Biological Sciences, Stanford University, Stanford, CA 94305.

from U.S. Biochemical Corp., and the vector pUCD2 (8) was obtained from T. Close.

The *virE* operon was subcloned as a 3.2-base-pair (bp) *Bam*HI fragment from pSW108 (34) into the *Bam*HI-cut phagemid pTZ18R. The recombinant plasmid with the 5′ end of the *virE* operon adjacent to the *Kpn*I site was designated pSG675. Plasmid pRS2003 was obtained by site-directed mutagenesis of pSG675 by the method of Taylor et al. (31). Single-stranded DNA was prepared from TG1 cells that were infected with helper phage M13KO7. The 35-mer oligonucleotide dCATTTATTGCCTGCTCAT/GCGATATCCGTTTC AAG (the slash indicates the fusion point) was used to delete the *vir* box that was present in the *virE* gene of pSG675. The DNA sequence at the fusion junction of the resulting pRS2003 was confirmed by sequence analysis (28). A custom-synthesized 18-mer oligonucleotide (dCCAGTTCCTCC CTTTCAG) complementary to the *virE* promoter region was used as a primer. The *virE* operon that was missing the *vir* box was subcloned as a *Pst*I Asp-718 fragment, after the 5′ sticky end of the Asp-718 site was filled in with Klenow fragment of DNA polymerase, into a *Pst*I-*Eco*RV restriction fragment of pUCD2, giving rise to pRS2412. The corresponding wild-type fragment was subcloned from pSG675, resulting in pRS2502.

Plasmid pSW174, which contained the *virG*::*lacZ* translational fusion in a broad-host range plasmid, has been described previously (37). Plasmid pSW167 was constructed by cloning a 1.5-kilobase-pair *Bgl*II-*Pst*I fragment containing the *virG* gene into a *Bam*HI-*Pst*I-cut pTZ18R vector. Plasmid pPC401, which coded for a native-sized VirG protein, was derived from pSW167 by fusing the ATG codon of *lacZ′* to the second *virG* codon AAA by the site-directed mutagenesis technique described above. Plasmid pSG662 was constructed by cloning a 2.2-kilobase *Sal*I fragment of the *virE* gene into the *Sal*I site of pTZ19R. The *virE* gene was oriented in the opposite direction to the *lacZ′* gene of the vector. Plasmid pSG665 was constructed by cloning a 0.4-kilobase *Hin*cII fragment containing the *virB* promoter into a *Hin*cII-cut pTZ19R in the same direction as the *lacZ′* gene of the vector.

Construction of *Bal* 31 deletion mutants. Plasmid pSG675 was linearized at its unique *Kpn*I site and treated with the exonuclease *Bal* 31. DNA (80 μg/ml) was incubated with 1.5 U of *Bal* 31 per ml at 37°C. Samples were removed at 3-min intervals from 9 to 30 min, and after phenol extraction, the DNA was ethanol precipitated. The DNA was cut with *Pst*I and used for cloning the DNA into the broad-host-range vector pUCD2. An *Eco*RV-*Pst*I restriction fragment of pUCD2, which was missing both the promoter and N-terminal region of the β-lactamase gene as well as the tetracycline resistance gene, was isolated and ligated with the *Pst*I-cut, *Bal* 31-treated pSG675. The deletion mutants were characterized initially by restriction enzyme analysis. Deletion endpoints of selected mutants were determined by DNA sequence analysis on a double-stranded DNA template (28). A synthetic 25-mer oligonucleotide (dGAAATTGC ATCAACGCATATAGCGC) complementary to pBR322-derived vector sequences 20 bp upstream of the *Eco*RV site was used as a sequencing primer.

Purification of the VirG–β-galactosidase fusion protein. *A. tumefaciens* A348(pSW174) was induced with 100 μM acetosyringone in induction medium as described previously (37). Twenty liters of induced cells was centrifuged and suspended in 100 ml of buffer A (20 mM Tris hydrochloride, 10 mM $MgCl_2$, 1.6 M NaCl, 10 mM β-mercaptoethanol [pH 7.4]). Cells were disrupted by two passages through a French pressure cell at 20,000 lb/in^2. Insoluble materials were removed by centrifuging the cells at 27,000 × *g* for 1 h. The cell extract was then diluted to 15 mg of protein per ml with buffer A and passed through a 5-ml bed volume of a β-galactosidase affinity column (*p*-aminobenzyl 1-thio-β-D-galactopyranoside-agarose) (33) that was equilibrated with the same buffer. The column was washed with 100 ml of buffer A, and the retained proteins were eluted with buffer B (100 mM sodium borate, 100 mM β-mercaptoethanol [pH 10]). β-Galactosidase activity was measured by an *o*-nitrophenyl-β-D-galactopyranoside assay (37) to monitor the protein elution pattern.

N-terminal amino acid sequencing. Purified protein from the affinity column was concentrated with a Centricon filter, electrophoresed on a sodium dodecyl sulfate (SDS)–7% polyacrylamide gel (14), and electroblotted onto an Immobilon membrane. The protein band was visualized by Coomassie brilliant blue staining, and the fusion protein band was excised to sequence its N-terminal amino acids as described by Matsudaira (19).

VirG protein purification and renaturation. *E. coli* DH5α(pPC401) was induced with 1 mM isopropyl-β-D-thiogalactopyranoside, for 3 h, and the insoluble pellet fraction was used to purify the VirG protein by the method of Hager and Burgess (11). The purified protein was suspended in buffer D (50 mM Tris hydrochloride [pH 8.0], 1 mM EDTA, 5 mM dithiothreitol, 6 M guanidine hydrochloride) to a final concentration of 0.05 mg/ml. This solution was allowed to dialyze against 200 volumes of renaturation buffer R (20 mM HEPES [*N*-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; pH 7.4], 50 mM KCl, 5 mM dithiothreitol, 50% glycerol) for 18 h at 4°C without stirring and then for an additional 18 h in fresh renaturation buffer with stirring. The renatured protein was stored at −20°C.

DNA-binding assay. A 300-bp *Eco*RI-*Ava*II fragment containing the *virE* promoter region (see Fig. 5A) was isolated from pSG662. [α-^{32}P]dATP and [α-^{32}P]dGTP were used to label the coding and noncoding strands, respectively, by the Klenow enzyme reaction. For the *virB* promoter, a 110-bp *Eco*RI-*Ava*II fragment (see Fig. 5A) was isolated from pSG665, and [α-^{32}P]dGTP and [α-^{32}P]dATP were used to label the coding and noncoding strands, respectively, by the Klenow enzyme reaction. For gel retardation assays, about 1 ng of a radiolabeled DNA fragment (10,000 cpm) was incubated with 0.2 μg of VirG protein and 1 μg of competitor DNA [either poly(dI-dC)(dI-dC) or calf thymus DNA] in a 10-μl binding buffer (20 mM HEPES, 50 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol [pH 7.4]. The mixture was incubated at room temperature for 20 min and then loaded onto a 5% nondenaturing polyacrylamide gel. The gel was electrophoresed in 0.25× TBE buffer (18) containing 2% glycerol at 15 V/cm. After electrophoresis, the gel was dried and exposed to AR-type X-ray film for 12 h at −70°C with an intensifying screen.

DNA footprinting assay. After the DNA-binding reaction described above proceeded for 20 min, 0.1 volume of 50 mM $CaCl_2$ and 100 mM $MgCl_2$ was added to the mixture followed by the addition of 0.5 ng of DNase I. After 1 min of incubation at room temperature, the mixture was loaded onto a 5% polyacrylamide gel and run at 15 V/cm for 2 h. The radioactive band was located by a brief exposure to AR-type X-ray film, and the retarded band was excised and electroeluted. The eluted DNA was precipitated with ethanol and then electrophoresed on a 6% DNA sequencing gel.

Western blot (immunoblot) analysis. *A. tumefaciens* was induced with acetosyringone essentially as described by

Winans et al. (37). *Agrobacterium tumefaciens* cells grown in induction broth with or without acetosyringone were collected by centrifugation and suspended in 1/25 of the original volume in 50 mM Tris hydrochloride (pH 7.4)–50 mM KCl–5 mM $MgCl_2$–1 mM dithiothreitol. Cells were disrupted by two passages through a French pressure cell at 20,000 lb/in^2. The protein concentration was determined by the method of Bradford (5). A total of 10 μg of protein of crude extracts was separated on an SDS–10% polyacrylamide gel as described by Laemmli (14). Proteins were transferred onto a nitrocellulose membrane and immunostained by using polyclonal anti-VirE2 antisera and alkaline phosphatase-conjugated goat anti-rabbit antisera.

RESULTS

The *virA* and *virG* genes have been shown to be required for the expression of all virulence genes of *A. tumefaciens* in response to plant phenolic compounds such as acetosyringone (30). According to a model proposed by Winans et al. (35), the VirA protein receives a signal from the plant and converts the VirG protein to an active form. The VirG protein activates *vir* gene transcription by physically binding to *vir* promoters. To provide experimental data to support this model, we first examined the 5′-nontranscribed region of the *virE* operon for regulatory sequences that may function as a putative VirG-binding site.

Minimum sequences necessary for *virE* induction. It has been shown previously (29, 34) that the fragment of 194 bp upstream of the transcriptional start site (position +1 according to Das et al. [9]) is sufficient for acetosyringone-dependent induction of the *virE* operon. To determine the minimum sequence necessary for the specific *virE* expression, a series of promoter deletion mutants was constructed. By using the exonuclease *Bal* 31, various portions of the 5′-noncoding sequences were removed, and the resulting fragments were cloned into the broad-host-range vector pUCD2. pBR322-derived promoter sequences of antibiotic resistance genes which could interfere with *virE* expression were removed as described in the Materials and Methods.

Representative deletions selected for further analysis occurred in the regions between the transcriptional start site and 224 bp of upstream sequences (Fig. 1A). The corresponding plasmid designations and the deletion endpoints, which were determined by DNA sequence analysis, are indicated in Fig. 1A. To analyze the acetosyringone-dependent induction of the *virE* operon, each deletion was introduced into a *virE* mutant of *Agrobacterium* strain Mx358, which contains a Tn*3*HoHo1 insertion in the *virE2* gene (29). This mutation resulted in the production of a VirE2-LacZ fusion protein. Plasmids were transferred from *E. coli* to *A. tumefaciens* by triparental mating (10) by using the helper strain HB101(pRK2013) for mobilization.

The expression of the *virE* operon was then studied by Western blot (immunoblot) analysis of total protein by using polyclonal anti-VirE2 antisera. Figure 1B shows that deletions up to position −70 with respect to the transcriptional start site (mutants pDE20, −34, −33, −04, −02, and −40) expressed wild-type levels of *virE2* (pDE31) in the presence of acetosyringone. Deletions which extended further downstream (mutants pDE38 and pDE19) completely abolished the induction of the *virE* operon. The VirE-LacZ fusion protein was detected in all lanes in Fig. 1B, thus providing a positive control of *vir* gene induction.

We concluded that 70 bp upstream of the transcriptional start site was sufficient for specific *virE* gene induction.

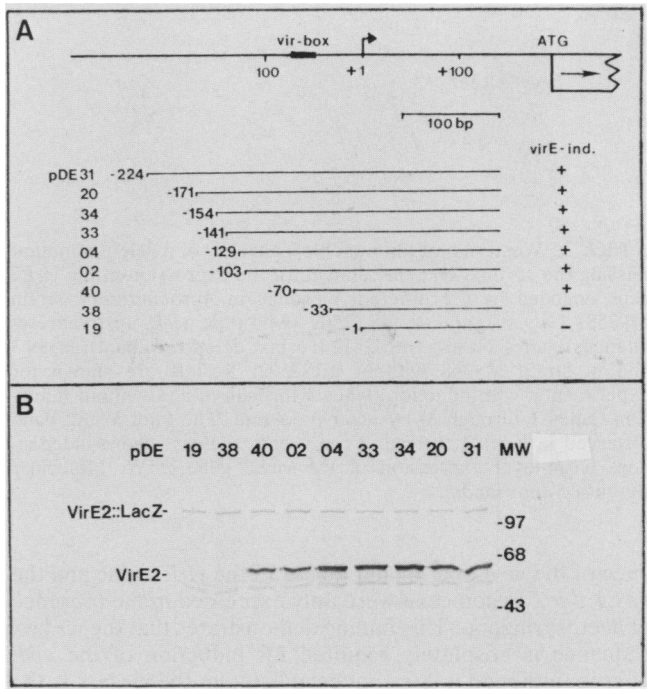

FIG. 1. Schematic representation of the noncoding sequences of the *virE* operon and the *Bal* 31 deletion mutants (A) and Western blot (immunoblot) analysis of the *Bal* 31 deletion mutants (B). (A) The arrow marks the transcriptional start site (position +1), and ATG represents the translational start of the *virE1* gene. The left column lists the plasmids with the various deletions in the 5′-noncoding sequences, and the right column shows the results of the Western blot analysis. The deletion endpoints are indicated with regard to the transcriptional start site. (B) The Western blot shows the expression of the *virE2* gene encoded by the various deletion mutants in *Agrobacterium* strain Mx358 after induction with acetosyringone. MW, Molecular weights (in thousands).

Since the regulatory sequences are located upstream of the putative −10 region (a conserved −35 region is not found in *Agrobacterium vir* genes) (9), they contain *cis*-acting upstream regulatory sequences.

Deletion of the *vir* box sequence. Based on DNA sequence analysis, conserved sequences have been identified in the 5′-nontranscribed regions of all *vir* genes (9, 36). The dodecadeoxynucleotide consensus sequence TNCAATTGAA APy, which was present as single or multiple copies, was postulated to play a role in *vir* gene regulation (36) and is designated the *vir* box. Recently, *vir* boxes have also been identified in the upstream of promoter regions of two other plant-inducible *Agrobacterium* genes, *virH* (*pinF*, [13]) and *tzs* (26), which are found on octopine or nopaline Ti plasmids, respectively. In the nontranscribed region of the *virE* operon (Fig. 1A), one *vir* box was located at positions −55 to −67, which was the 5′ end of the minimal sequences required for *virE* induction (position −70; pDE40).

To investigate whether the *vir* box sequence was involved in transcriptional activation, we deleted this 12-bp sequence (TGCAGTTGAAAC) by site-directed mutagenesis. The transcriptional activation of the *virE* operon was then analyzed by Western blot analysis as described above for the *Bal* 31 deletion mutants. Figure 2 shows that acetosyringone-dependent induction of the *virE* gene was completely abolished when the *vir* box was deleted. Control experiments demonstrated that the VirE2 signal depended on the pres-

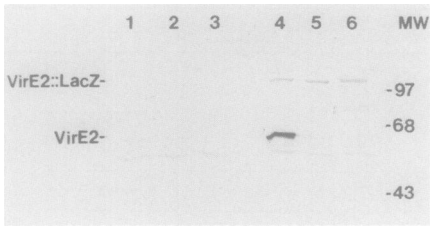

FIG. 2. Western blot (immunoblot) analysis of a deletion mutant missing the *vir* box. Data are shown for the expression of the *virE2* gene encoded by the different plasmids in *Agrobacterium* strain Mx358. Lanes 1 and 4, pRS2502 (wild-type *virE* upstream sequences); lanes 2 and 5, pRS2412 (*vir* box deletion mutant); lanes 3 and 6, strain Mx358 without pRS2502. Results are shown for experiments with induction (lanes 4 through 6) and without induction (lanes 1 through 3) by acetosyringone. The faint VirE2 band observed in lanes 1 through 3 represents acetosyringone-independent basal-level expression of the *virE2* gene. MW, Molecular weight (in thousands).

ence of the *virE* operon and that both the *virE2* gene and the *virE2*::*lacZ* fusion gene were only expressed in the presence of acetosyringone. This finding demonstrates that the *vir* box sequence is absolutely required for induction of the *virE* operon, although it does not establish that the *vir* box is the only *cis*-acting regulatory sequence present upstream of the promoter region.

Once we showed that the *vir* box was essential for induction, we determined whether the VirG protein bound specifically to the *vir* box. The promoter regions from two *vir* genes were studied: *virB*, which has two *vir* boxes, and *virE*, which has only one *vir* box.

Determination of the *virG* translational start site. Since we were unable to overproduce the VirG protein in *E. coli* by transcriptional fusion using the *tac* promoter of pMY1133 (39), it was necessary to determine the translational start site in order to overproduce the native-sized VirG protein. To do so, we sequenced the N-terminal amino acids of the VirG–β-galactosidase fusion protein encoded by pSW174 (37). *A. tumefaciens* A348 containing pSW174 was induced with acetosyringone, and the soluble cell extract was passed through an affinity column against β-galactosidase protein (*p*-aminobenzyl 1-thio-β-D-galactopyranoside-agarose). Bound proteins were eluted with a high-pH solution as described in the Materials and Methods. About 2 mg of bound protein was obtained from 20 liters of induced cells. When the eluant was electrophoresed on a SDS–7% polyacrylamide gel and stained with Coomassie brilliant blue, two high-molecular-weight protein bands were visualized (Fig. 3A). Antibody against β-galactosidase recognized both bands (Fig. 3B), while only the higher-molecular-weight band was recognized by antibody against VirG (data not shown). This suggests that the higher-molecular-weight band is the VirG–β-galactosidase fusion protein, while the lower-molecular-weight band is β-galactosidase, which was probably derived from cleavage of the fusion protein. Since the VirG portion of the fusion protein had a molecular size of 22 kilodaltons (kDa) and the β-galactosidase portion had a molecular size of 116 kDa, the estimated apparent molecular sizes of 120 and 140 kDa on the SDS-polyacrylamide gel further supported this conclusion.

The eluted protein was concentrated, and the N-terminal end of the VirG–β-galactosidase fusion protein was sequenced as described in the Materials and Methods. The N-terminal amino acid sequence of the fusion protein was

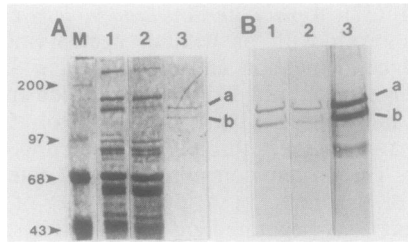

FIG. 3. Purification of VirG–β-galactosidase fusion protein. (A) Coomassie brilliant blue stain. (B) Immunoblot with antibody against β-galactosidase. Lanes 1, Soluble cell extract of the acetosyringone-induced A348(pSW174); lanes 2, pass through of the affinity column (*p*-aminobenzyl 1-thio-β-D-galactopyranoside-agarose); lanes 3, eluant from the affinity column; M, molecular weight standards (indicated to the left of the gel, in thousands); a, VirG–β-galactosidase fusion protein; b, β-galactosidase.

determined to be MKHVLLV, which corresponds to a TTG start site 26 codons downstream from the previously predicted ATG start site (35).

Overproduction of native-sized VirG protein. By using the data obtained above, a plasmid coding for a native-sized VirG protein was constructed by a site-directed mutagenesis method described previously (31). An oligonucleotide, 5′-dCAGGAAACAGCTATGAAACACGTTCTTCTT-3′, was used to delete the sequence between the second *lacZ′* codon ACC and the first *virG* codon TTG. The resulting construct, pPC401, which was confirmed by sequence analysis, fused the first codon ATG of the *lacZ′* gene with the second codon AAA of the *virG* gene (Fig. 4A). Plasmid pPC401-encoded native sized VirG protein was insoluble in *E. coli* when it was induced with isopropyl-β-D-thiogalactopyranoside and had an apparent molecular size of 25 kDa on SDS-polyacrylamide gel electrophoresis (Fig. 4B), which was close to its predicted size of 27 kDa. The native VirG protein that was produced in *A. tumefaciens*, which was detected by an immunoblot assay by using antibody against VirG, was the same size as the VirG protein coded by pPC401 (data not shown), which further confirms that TTG is the translational start site.

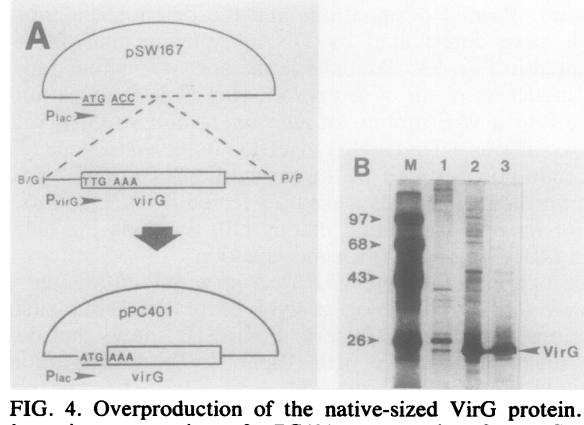

FIG. 4. Overproduction of the native-sized VirG protein. (A) Schematic presentation of pPC401 construction from pSW167. Plasmid pPC401 encoded a native-sized VirG protein under the control of the *lacZ* promoter. B/G, *Bam*HI-*Bgl*II; P/P, *Pst*I-*Pst*I. (B) Coomassie brilliant blue staining; lane 1, insoluble proteins from isopropyl-β-D-thiogalactopyranoside-induced DH5α(pSW167); lane 2, insoluble proteins from isopropyl-β-D-thiogalactopyranoside-induced DH5α(pPC401); lane 3, gel-purified VirG protein; M, molecular weight standards (indicated to the left of the gel, in thousands).

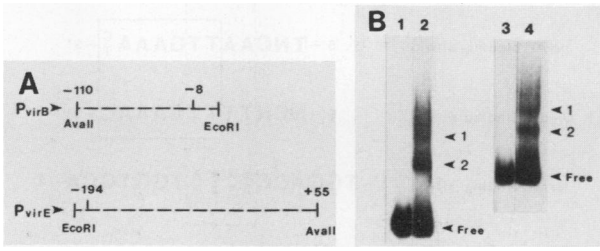

FIG. 5. Gel retardation assay with both the *virB* and *virE* promoter regions. (A) DNA fragments used in the assay. The numbers indicate the position relative to the inducible transcriptional start site. (B) Gel retardation assay with *virB* (lanes 1 and 2) and *virE* (lanes 3 and 4) in the presence (lanes 2 and 4) or absence (lanes 1 and 3) of the VirG protein. In each case, a 1,000 times excess of poly(dI-dC)(dI-dC) was used (see text for details of the assay). The positions of the two retarded bands are indicated by numbers.

VirG protein specifically binds to the *vir* box. The pPC401-encoded VirG protein was purified and renatured as described in the Materials and Methods. Using this protein, we conducted DNA-binding assays. In gel retardation assays, the VirG protein specifically retarded DNA fragments containing either the *virE* or *virB* promoters in the presence of a 1,000 times excess of competitor DNA (Fig. 5). In both cases, two species of retarded bands were observed. Under the same conditions, no retardation was observed when the labeled DNA fragment was from either the *virE* or *virB* structural genes (data not shown).

In order to determine the VirG-binding sequence, we carried out a footprinting analysis first on the *virE* promoter. Figure 6A and B shows footprinting results of the fast-migrating retarded band (band 2 of lane 4 in Fig. 5B). The VirG protein specifically protected 22 nucleotides on the coding and 37 nucleotides on the noncoding strands of the *virE* promoter. It is clear from Fig. 6A and B that the *vir* box is in the center of the protected area. This result suggests that the VirG protein recognizes the *vir* box on the *virE* promoter.

The slow-migrating retarded band was also used to conduct the footprinting analysis. Comparing the footprinting results of the two retarded species, we found no difference in the footprinting patterns (data not shown). It is not clear why two retarded bands formed and had the same DNase I protection patterns, but this has also been observed in another system (15).

To further confirm the suggestion that the VirG protein recognizes the *vir* box, we conducted footprinting assays on the *virB* promoter, which has two *vir* boxes. As shown in Fig. 6C and D, 74 nucleotides on the coding strand and 48 nucleotides on the noncoding strand were protected from DNase I digestion in the fast-migrating retarded band (band 2 of lane 2 in Fig. 5B). The protected area covered both of the *vir* boxes, and the two *vir* boxes were located in the center of the protected area. This confirmed our prediction that the VirG protein recognizes the *vir* box. However, the first *vir* box (positions −68 and −57) was protected much more strongly than was the second *vir* box on both strands (Fig. 6C and D), since some nucleotide bands were observable in the region corresponding to the second *vir* box but not the first one. This comparison of the strength of binding was even more striking when the film was overexposed (data not shown). These data suggest that the first *vir* box may be more important than the second one in *virB* gene induction. Like the *virE* promoter, both species of retarded bands gave rise to the same DNase I protection patterns.

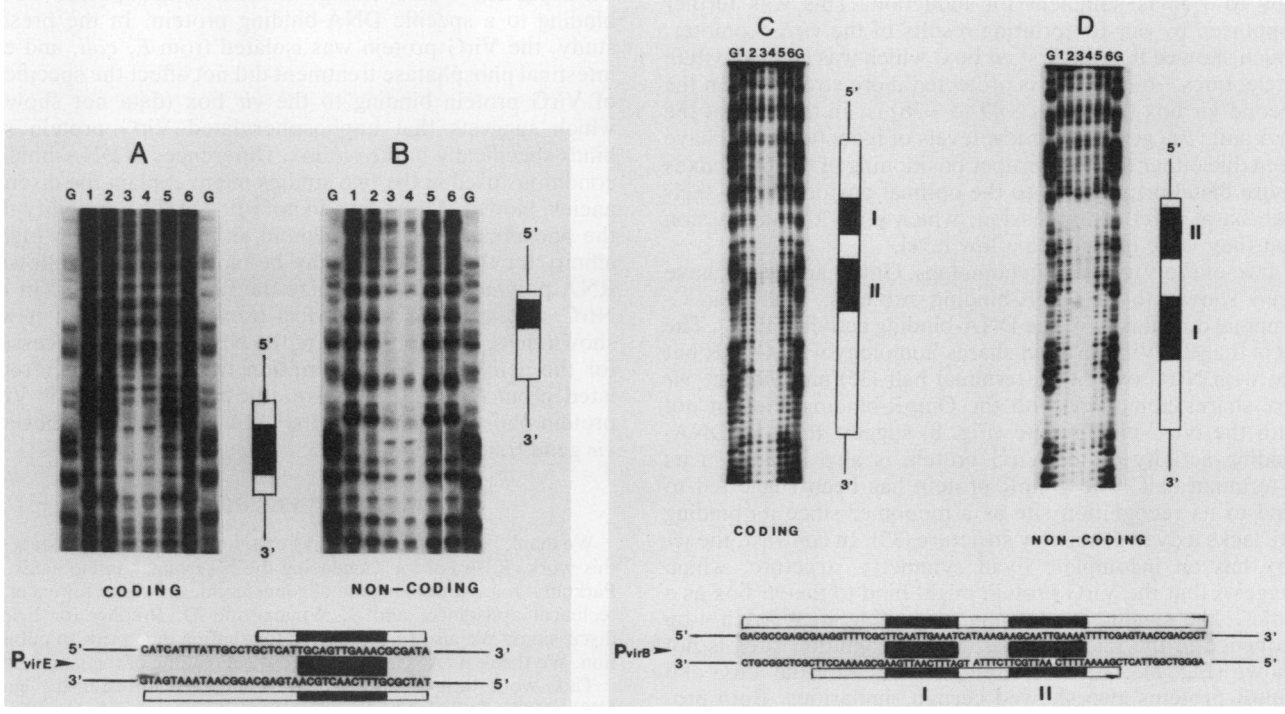

FIG. 6. DNase I footprinting analysis of the *virE* and *virB* promoter regions. (A) Coding strand fo the *virE* promoter; (B) noncoding strand of the *virE* promoter region; (C) coding strand of the *virB* promoter region; (D) noncoding strnad of the *virB* promoter region. Lane G, Guanine chemical cleavage (20); lanes 1, 2, 5, and 6 DNase I digest without VirG bound; lanes 3 and 4, DNase I digest with VirG bound (see text for details of the assay). The regions enclosed in the open boxes are the protected areas and the black boxes indicate the *vir* box.

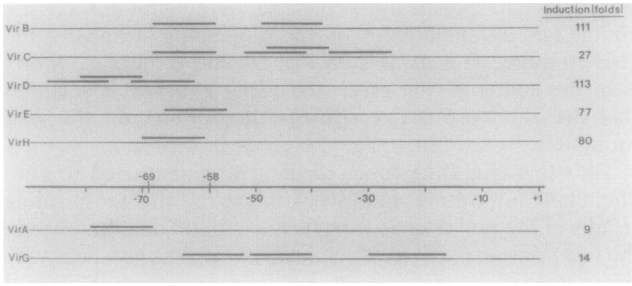

FIG. 7. The *vir* box positions in different *vir* gene promoters and the inducibility of each gene (13, 30, 38). Dark lines represent the *vir* box, and +1 indicates the inducible transcriptional start site.

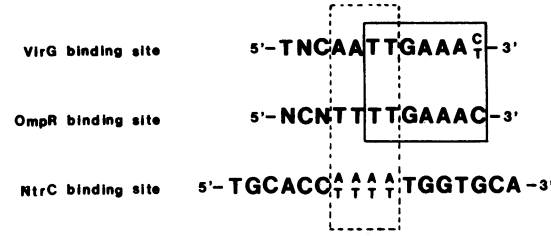

FIG. 8. Comparison of DNA-binding sequences of three homologous proteins (1, 22). The solid box indicates homologous regions of the VirG and OmpR-binding sites; the dotted box indicates an AT-rich region found in all three binding sequences.

DISCUSSION

During the past several years, upstream activation sequences have been identified in the 5'-flanking sequences of several procaryotic genes (4, 24), and a few of these elements function in a way similar to eucaryotic enhancer elements (2, 21). Regulatory elements that are functionally and structurally similar to the *Agrobacterium vir* box have been identified in other bacterial systems that are involved in signal transduction (17, 24).

Analysis of the different *vir* gene promoters suggests that not only the presence of the *vir* box but also its relative position are important in the regulation of transcription. As shown in Fig. 7, all highly inducible *vir* genes, *virB*, *-C*, *-D*, *-E*, and *-H* (*pinF*) (27- to 113-fold) (13, 29), have at least one *vir* box located at the same position (positions −69 to −58, ±3bp) relative to the inducible transcriptional start site. Among them, the *virE* and *virH* genes have only one copy of the *vir* box, which suggests that only one *vir* box at positions −69 to −58 is sufficient for induction. This was further supported by our footprinting results of the *virB* promoter, which showed that the first *vir* box, which was located within nucleotides −69 to −58, is protected more strongly than the second vir box (positions −49 to −38) is. In the case of the *virA* and *virG* genes, their low levels of induction could have been due either to the improper positioning of their *vir* boxes (more distal or proximal to the optimal position) or to their high basal level of expression, which gives the impression that they were induced to a low level.

Two of the VirG protein homologs, OmpR and NtrC, have been shown to be DNA-binding proteins, and their C-terminal domains have the DNA-binding activity (1, 32). The facts that the VirG protein shares homology with OmpR but not with NtrC over its C-terminal half (35) and that the *vir* box shares homology with the OmpR-binding site but not with the NtrC-binding site (Fig. 8) suggest that the DNA-binding activity of the VirG protein is also located in its C-terminal half. The OmpR protein has been suggested to bind to its recognition site as a monomer since its binding site lacks a dyad symmetry structure (32). In contrast, the *vir* box has an incomplete dyad symmetry structure, which suggests that the VirG protein might bind to the *vir* box as a dimer. The significance, if any, of having an AT-rich 4-bp sequence in the middle of all three recognition sites is not known (Fig. 8). The footprinting results with the VirG and OmpR proteins also showed certain similarities. Both proteins protect long stretches of DNA: more than 50 nucleotides of the *ompF* promoter by OmpR (22) and 37 nucleotides of the *virE* promoter by VirG. However, both recognition sites have a 12-nucleotide sequence in common. This suggests that the proteins may bind to DNA cooperatively in both systems.

The fact that the *virG* gene used the TTG codon and not an upstream ATG sequence as a start site is consistent with several observations which were rather difficult to reconcile with the fact that ATG served as the start site. First, the TTG codon is preceded by a reasonable ribosome-binding site, while the upstream ATG codon is not (35); second, the TTG codon is located at a reasonable distance from both inducible and constitutive transcriptional start sites, while the ATG codon resides only 1 bp downstream of the constitutive transcriptional start site (30); third, the TTG codon resides at approximately the same position as the translational start sites of its homologs (35).

We have recently shown that the VirA protein is autophosphorylated (12) and that it can also phosphorylate the VirG protein in vitro (S. Jin, T. Roitsch, and E. W. Nester, manuscript in preparation). In a recent study, Powell et al. (25) discussed the possibility that phosphorylation of the VirG protein would change it from a nonspecific DNA-binding to a specific DNA-binding protein. In the present study, the VirG protein was isolated from *E. coli*, and calf intestinal phosphatase treatment did not affect the specificity of VirG protein binding to the *vir* box (data not shown), which suggests that unphosphorylated VirG protein still binds specifically to the *vir* box. Differences in DNA-binding conditions used in the two studies might explain the discrepancies. However, our data do not rule out the possibility that the phosphorylated VirG protein either may have a higher affinity for the vir box or may be more efficient in allowing RNA polymerase to recognize the promoter region. In the NtrC system, using an in vitro transcription assay, it was shown that phosphorylation of the NtrC protein is necessary for the activation of transcription from the nitrogen-regulated promoters (23). We are studying the effect of VirG protein phosphorylation on its DNA-binding properties and *vir* gene transcription.

ACKNOWLEDGMENTS

We thank W. Sim for conducting experiments in the early stages of this work, R. Prusti for sequencing the N-terminal amino acids, D. Parkhurst for synthesizing the oligonucleotides, P. Lam for excellent technical assistance, and S. Winans and D. Bradley for helpful discussions. We also thank A. Das for sharing data prior to publication. We thank R. Ankenbauer for critical reading of the manuscript.

This work was supported by Public Health Service grant 5R01GM32618-17 from the National Institutes of Health (to E.W.N.). T.R. was supported by postdoctoral fellowship RO758/1-1 from the Deutsche Forschungsgemeinschaft.

S.J. and T.R. made equal contributions to this report; these authors are listed in alphabetical order.

ADDENDUM IN PROOF

A. Das and G. Pazour have recently shown that the *vir* box is important for acetosyringone-mediated induction of the *virB* gene (Nucleic Acids Res. **17**:4541–4550, 1989).

LITERATURE CITED

1. **Ames, G., and K. Nidaido.** 1985. Nitrogen regulation in *Salmonella typhimurium*. Identification of an *ntrC* protein-binding site and definition of a consensus binding sequence. EMBO J. **4**:539–547.
2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β globin gene enhanced by remote SV 40 DNA sequences. Cell **27**:299–308.
3. **Binns, A. N., and M. F. Thomashow.** 1988. Cell biology of *Agrobacterium* infection and transformation of plants. Annu. Rev. Microbiol. **42**:1–35.
4. **Birkmann, A., and A. Bock.** 1989. Characterization of a cis acting regulatory DNA element necessary for formate induction of the formate dehydrogenase gene (fdhF) of *Escherichia coli*. Mol. Microbiol. **3**:187–195.
5. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
6. **Christie, P. J., J. E. Ward, S. C. Winans, and E. W. Nester.** 1988. The *Agrobacterium tumefaciens virE2* gene product is a single-stranded DNA-binding protein that associates with T-DNA. J. Bacteriol. **170**:2659–2667.
7. **Citovsky, V., G. De Vos, and P. Zambryski.** 1988. Single strand DNA binding protein encoded by the virE locus of *Agrobacterium* tumefaciens. Science **240**:501–506.
8. **Close, T. J., D. Zaitlin, and C. I. Kado.** 1984. Design and development of amplifiable broad-host-range cloning vectors: analysis of the *vir* region of *Agrobacterium tumefaciens* plasmid pTiC58. Plasmid **12**:111–118.
9. **Das, A., S. Stachel, P. Ebert, P. Allenza, A. Montoya, and E. W. Nester.** 1986. Promoters of *Agrobacterium tumefaciens* Ti-plasmid virulence genes. Nucleic Acids Res. **14**:1355–1364.
10. **Ditta, G., S. Stanfield, D. Corbin, and D. Helinski.** 1980. Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. USA **77**:7347–7351.
11. **Hager, D., and R. Burgess.** 1980. Elution of protein from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. Anal. Biochem. **109**:76–86.
12. **Jin, S., T. Roitsch, R. G. Ankenbauer, M. P. Gordon, and E. W. Nester.** 1990. The VirA protein of *Agrobacterium tumefaciens* is autophosphorylated and is essential for *vir* gene regulation. J. Bacteriol. **172**:525–530.
13. **Kanemoto, R., A. Powell, D. Akiyoshi, D. Regier, R. Kerstetter, E. Nester, M. Hawes, and M. Gordon.** 1989. Nucleotide sequence and analysis of the plant-inducible locus *pinF* from *Agrobacterium tumefaciens*. J. Bacteriol. **171**:2506–2512.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (london) **227**:680–685.
15. **Landolfi, N., J. Capra, and P. Tucker.** 1987. Protein-nucleotide contacts in the immunoglobulin heavy-chain promoter region. Proc. Natl. Acad. Sci. USA **84**:3851–3855.
16. **Leroux, B., M. F. Yanofsky, S. C. Winans, J. E. Ward, S. F. Zeigler, and E. W. Nester.** 1987. Characterization of the *virA* locus of *Agrobacterium tumefaciens*: a transcriptional regulator and host range determinant. EMBO J. **6**:849–856.
17. **Maeda, S., Y. Ozawa, T. Mizuno, and S. Mizushima.** 1988. Stereospecific positioning of the cis-acting sequence with respect to the canonical promoter is required for activation of the *ompC* gene by a positive regulator, OmpR, in *Escherichia coli*. J. Mol. Biol. **202**:433–441.
18. **Maniatis, T., E. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 156. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262**:10035–10038.
20. **Maxam, A., and W. Gilbert.** 1980. Sequencing with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
22. **Mizuno, T., M. Kato, Y.-L. Jo, and S. Mizushima.** 1988. Interaction of OmpR, a positive regulator, with the osmoregulated *ompC* and *ompF* genes of *Escherichia coli*. J. Biol. Chem. **263**:1008–1012.
23. **Ninfa, A., and B. Magasanik.** 1986. Covalent modification of the *glnG* product, NRI, by the *glnL* product, NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli*.Proc. Natl. Acad. Sci. USA **83**:5909–5913.
24. **Popham, D. L., D. Szeto, J. Keener, and S. Kustu.** 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. Science **243**:629–635.
25. **Powell, B., P. Rogowsky, and C. Kado.** 1989. *virG* of *Agrobacterium tumefaciens* plasmid pTiC58 encodes a DNA-binding protein. Mol. Microbiol. **3**:411–419.
26. **Powell, G. K., N. G. Hommes, J. Kuo, L. A Castle, and R. O. Morris.** 1988. Inducible expression of cytokinin biosynthesis in *Agrobacterium tumefaciens* by plant phenolics. Mol. Plant-Microbe Interact. **1**:235–242.
27. **Ronson, C. W., B. T. Nixon, and F. M. Asubel.** 1987. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell **49**:579–581.
28. **Sanger, F., S. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Stachel, S., and E. Nester.** 1986. The genetic and transcriptional organization of the *vir* region of the A6 Ti plasmid of *Agrobacterium tumefaciens*. EMBO J. **5**:1445–1454.
30. **Stachel, S., and P. Zambryski.** 1986. VirA and VirG control the plant-inducible activation of the T-DNA transfer process of *A. tumefaciens*. Cell **46**:325–333.
31. **Taylor, J., J. Ott, and F. Eckstein.** 1985. The rapid generation of oligonucleotide-directed mutagenesis at high frequency using phosphorothioate-modified DNA. Nucleic Acids Res. **13**:8765–8785.
32. **Tsung, K., R. Brissette, and M. Inouye.** 1989. Identification of the DNA binding domain of the OmpR protein required for transcriptional activation of the *ompF* and *ompC* genes of *Escherichia coli* by *in vivo* DNA footprinting. J. Biol. Chem. **264**:10104–10109.
33. **Ullmann, A.** 1984. One-step purification of hybrid proteins which have β-galactosidase activity. Gene **29**:27–31.
34. **Winans, S., P. Allenza, S. Stachel, K. McBride, and E. Nester.** 1987. Characterization of the *virE* operon of the Agrobacterium Ti plasmid pTiA6. Nucleic Acids Res. **15**:825–837.
35. **Winans, S., P. Ebert, S. Stachel, M. Gordon, and E. Nester.** 1986. A gene essential for *Agrobacterium* virulence is homologous to a family of positive regulatory loci. Proc. Natl. Acad. Sci. USA **83**:8278–8282.
36. **Winans, S., S. Jin, T. Komari, K. Johnson, and E. Nester.** 1987. The role of virulence regulatory loci in determining Agrobacterium host range, p. 573–582. *In* D. Von Wettstein and N. Chua (ed.), Plant molecular biology. Plenum Publishing Corp., New York.
37. **Winans, S., R. Kerstetter, and E. Nester.** 1988. Transcriptional regulation of the *virA* and *virG* genes of *Agrobacterium tumefaciens*. J. Bacteriol. **170**:4047–4054.
38. **Winans, S., R. Kerstetter, J. Ward, and E. Nester.** 1989. A protein required for transcriptional regulation of *Agrobacterium* virulence genes spans the cytoplasmic membrane. J. Bacteriol. **171**:1616–1622.
39. **Yanofsky, M., S. Porter, C. Young, L. Albright, M. Gordon, and E. Nester.** 1986. The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease. Cell **47**:471–477.
40. **Zambryski, P., J. Tempe, and J. Schell.** 1989. Transfer and function of T-DNA genes from *Agrobacterium* Ti-and Ri plasmids in plants. Cell **56**:193–201.