Characterization of the VirG binding site of *Agrobacterium tumefaciens*

Gregory J.Pazour and Anath Das*
Department of Biochemistry and Plant Molecular Genetics Institute, University of Minnesota, 1479 Gortner Avenue, St Paul, MN 55108, USA

Received September 5, 1990; Revised and Accepted November 5, 1990

ABSTRACT

Expression of *Agrobacterium tumefaciens* virulence (*vir*) genes is dependent on the presence of a conserved 'vir box' sequence in their 5' nontranscribed regions. The location and number of these sequences vary considerably in different *vir* genes. Site-directed mutagenesis was used to indentify the functional vir box(es) of *vir*B, *vir*C and *vir*D. For *vir*B expression both vir box B1 and B2 are required but only the vir box B1 is absolutely essential. Of the five vir boxes of *vir*C and *vir*D two are required for *vir*C expression while only one vir box is required for *vir*D expression. To investigate the minimum sequences necessary for vir gene induction a deletion derivative of *vir*E that lacks the vir box region was used. This mutant is not induced by acetosyringone. The inducibility of this promoter was restored when a synthetic deoxyoligonucleotide dGTTTCAATTGAAAC was introduced at a location analogous to that of the wild type vir box sequence. Mutational analysis indicate that the functional vir box sequence is 14 residues in length, contains a dyad symmetry and has the consensus sequence d ryTnAa TTGnAaY (r= purine, y= pyrimidine).

INTRODUCTION

Transcription of the virulence (*vir*) genes of the plant pathogen *Agrobacterium tumefaciens* is positively regulated by VirA and VirG (1,2,3). Polypeptides encoded by these two regulatory genes are highly homologous to other prokaryotic positive regulatory proteins that are members of two component regulatory systems (4,5,6). In two component regulatory systems one component, a histidine kinase, senses the environment and then signals the second, the response regulator, to carry out a function. Usually the response regulator affects transcription although it can have other functions (7). VirA, the histidine kinase of the *A. tumefaciens* virulence system (8,9) appears to sense the presence of plant cells by interacting with phenolic compounds, e.g. acetosyringone (AS), produced by susceptible plant cells (10). Such interaction is hypothesized to lead to the autophosphorylation of VirA (8,9) which in turn phosphorylates VirG, the response regulator (11). VirG is a sequence specific DNA binding protein which binds to DNA sequences located upstream of the virulence promoters and may function by activating transcription (12,13). The VirG binding site, the vir box, has the consensus 5'dPu(T/A)TDCAATTGHAAPy (H=A,C or T, D=A,G, or T) (14,15,16). Deletion and site-specific mutagenesis studies showed that one or more vir box sequences are required for expression of *vir*B, *vir*E and *vir*G (12,15,17). The number and location of the vir box sequence is highly variable in all the inducible *vir* loci of both the pTiA6 and pTiC58 plasmids. The vir box of the closely related Ri plasmid *vir* genes has the sequence 5'dTG(A/T)AA(C/T) (18) which is equivalent to the 3' half of the Ti plasmid vir box.

DNA binding and gene expression studies indicate that only a subset of the vir box sequences are required for expression of a specific *vir* gene. The *vir*G gene has three copies of the vir box sequence upstream of the start of transcription; two of these are bound by VirG *in vitro* (13). Functional studies of Winans (17) showed that these same two boxes bound by VirG *in vitro* are required for AS induced expression of *vir*G while the third one is not required for induction by AS. DNase I protection studies revealed that two of the five vir boxes between the divergent *vir*C and *vir*D operons are bound by VirG *in vitro*. In addition a highly AT rich region around −25 of the *vir*C operon was protected by VirG from DNase 1 digestion (13). Whether these regions are required for expression of *vir*C, *vir*D or both operons was not known before this study. The two vir boxes of *vir*B are bound by VirG (12). Of these, the upstream box (distal to the promoter) is required for *vir*B induction (15) while the role of the downstream box was unknown before this study. The *vir*E operon contains one vir box which is protected by VirG and is required for its expression (12). The remaining *vir* operons, *vir*A, *vir*F and *vir*H, each contain at least one vir box. The role of these boxes is unknown.

In the present study we used site-specific mutagenesis procedures to examine the role of specific vir boxes in *vir* gene expression. We demonstrate that induction of the *vir*D operon requires only one vir box while induction of the *vir*B and *vir*C operons involve two boxes−one of these is absolutely required and the second contributes but is not absolutely required. The

* To whom correspondence should be addressed

major box of *vir*C is the same box that is required for *vir*D expression. We also examined a series of changes to the *vir*E vir box and show that inducibility of a deleted *vir*E promoter can be restored by addition of an oligonucleotide that contains the sequence 5′ dGTTTCAATTGAAAC.

MATERIALS AND METHODS

*vir*C, *vir*D assay plasmids

The intergenic region of the pTiA6 *vir*C and *vir*D operons was cloned into the plasmid vector pUC118 (19) to construct pGP254.1. This plasmid contains the ~ 0.5kb BamHI to Xma III fragment (20) of pTiA6 BamHI fragment 25 flanked by BamHI sites. The BamHI fragment of pGP254.1 was recloned into the BamHI site of the wide host range *lac*Z translational fusion vector pAD1227 in both orientations; one orientation (pGP264c) yielded a *vir*Cl:*lac*Z translational fusion while the other (pGP264d) yielded a *vir*D1:*lac*Z translational fusion. pAD1227 was constructed by deletion of a 0.33 kb BamHI fragment of pAD1092 (15).

Site directed mutagenesis was carried out on single-stranded pGP254.1 DNA according to the method of Kunkel (21). Synthetic deoxynucleotides (35−36 mers) were used as mutagenic primers. The central hexameric core region (dCAATTG or its homologues) of each vir box was altered to dTCTAGA (an XbaI restriction site). In addition the sequence dATAAAA at positions −26 to −21 of *vir*C (transcription initiation site +1(23)), designated TAAA, was also changed to dTCTAGA. All mutants were recloned as BamHI fragments into the reporter plasmid pAD1227 to construct *vir*C:*lac*Z and *vir*D:*lac*Z fusion plasmids. The mutant plasmids and their site of mutations are as follows: pGP265c,pGP265d (vir box C1D5); pGP266.1c,pGP266.1d (vir box C2D4); pGP267c,pGP267d (vir box C3D3); pGP268c,pGP268d (vir box C4D2); pGP269c,pGP269d (vir box C5D1); pGP270c,pGP270d.1 (TAAA) (see Table 1A). Mutants pGP294c and pGP294d are derivatives of pGP267c and pGP267d respectively, where the dTCTAGA mutation of vir box C3D3 was altered to dAAAAGA. All mutants described in this study were confirmed by DNA sequence analysis (22).

*vir*B assay plasmids

The 5′ region of the *vir*B operon was cloned as a 0.33kb BamHI fragment from pAD1092 (15) into pUC118 (19) to constuct pAD1221. Site directed mutagenesis was performed on pAD1221 with synthetic deoxynucleotides to substitute the central four bases (dAATT) of each vir box to dCCGG. The mutagenized BamHI fragments were recloned into pAD1227 to yield pGP245 (vir box B1) and pGP244.7 (vir box B2).

*vir*E analysis plasmids

Plasmid pDS1c contains a *vir*E:*lac*Z translational fusion flanked by a unique EcoRI site upstream of *vir*E and a unique PstI site downstream of *lac*Z. This fusion contains the *vir*E region from ~ −500 to +425 (23,24) with respect to the start of *vir*E transcription. In this construct the 20th codon of *vir*E2 was fused to the 8th codon of *lac*Z.

The deletion mutation pΔ27c was constructed by Bal31 exonuclease treatment of *vir*E upstream sequences followed by ligation to an EcoRI linker DNA and recircularization. The deletion endpoint was determined by DNA sequence analysis. To avoid uncertainty in the vector sequences due to Bal31 treatment the EcoRI to PstI fragment encompassing the *vir*E:*lac*Z fusion was cloned into a new vector.

Plasmid pGP275, pGP322 and pGP323 contain mutations in the vir box E1. In these plasmids the central twelve bases of the vir box E1 were converted from dTTGCAGTTGAAA to dTTGTCTAGAAAA, dTTTCAATTGAAA and dTTTCAATTTAAA, respectively, by site directed mutagenesis. Plasmid pGP304 and pGP321 were constructed by linearization of pΔ27c with EcoRI followed by ligation with self annealing oligonucleotide adaptors 5′dAATTAGCTTTCAATTGAAAGCT and 5′dAATTAGGTTTCAATTGAAACCT, respectively.

Other procedures

Plasmids were mobilized from *E. coli* to *A. tumefaciens* A348 by a triparental mating procedure (25). *Agrobacteria* were grown overnight in AB media (26), pH 7.0 and diluted 1:20 into AB mes medium (1× AB salts (26), 0.2% glucose, 25 mM phosphate, 25 mM 2-(N-morpholino)ethanesulfonic acid [MES] [pH5.5]) (± 100 μM Acetosyringone) to monitor *vir* gene induction. Cells were grown for 24 hr at 28°C for induction assays. β galactosidase enzyme activity was assayed according to Miller (27).

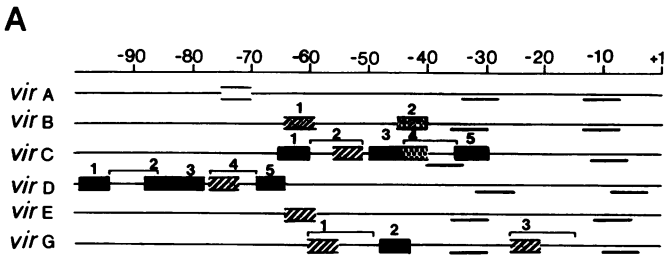
RESULTS

Vir Box C2D4 is required for both *vir*C and *vir*D expression while vir box C4D2 is necessary but not sufficient for *vir*C expression.

The divergent *vir*C and *vir*D operons share common 5′ nontranscribed regions. Within this region lie five copies of the vir box sequence. Because the vir box is palindromic individual boxes could potentially serve both operons or different boxes may be required for expression of each operon. DNase I protection studies showed that the transcriptional activator protein, VirG, binds to three sites within this region. Two of these correlate with the vir boxes C2D4 and C4D2 (see Figure 1A for numbering) while the other maps to a highly AT rich sequence which we have designated TAAA in Table 1. Only the two underlined A residues were protected in this region (13).

To determine the role of individual boxes in the regulation of *vir*C and *vir*D expression the intergenic region between these operons was cloned onto a wild host range *lac*Z translational fusion vector to yield a *vir*C:*lac*Z translational fusion in one orientation and a *vir*D:*lac*Z translational fusion in the opposite orientation. The central six bases (dCAATTG or a homologue) of each of the five vir boxes and the region encompassing the TAAA sequence were mutagenized to dTCTAGA. The effect of these alterations was measured by monitoring β-galactosidase activity from the *vir*C:*lac*Z or the *vir*D:*lac*Z fusion plasmid (Table 1).

Alteration of the vir box C2D4 (pGP266.1d) completely abolished induction of *vir*D while changes to the TAAA sequence (pGP270d) and C3D3 (pGP267d) attenuated induction as compared to wild type levels (pGP264d). Mutations in the other vir boxes had no effect on *vir*D expression (pGP265d, 268d, 269d). Vir boxes C2D4 and C3D3 are overlapping; the alteration of C3D3 changes the C2D4 vir box from dCTACAATTGCAATTT to dCTACAATTGTCTAGA (mutations are underlined thus it is plausible that the attenuated expression seen from the C3D3 mutant is due to disruption of C3D3, C2D4 or both. To distinguish between these possibilities the C3D3 mutant was further mutated to recreate the vir box C2D4 without recreating C3D3 (pGP294d). This alteration restored *vir*D



B

B1	G	C	T	T	C	A	A	T	T	G	A	A	A	T
E	A	T	T	G	C	A	G	T	T	G	A	A	A	C
G1	G	T	T	T	C	A	C	T	T	G	T	A	A	C
G3	C	A	T	A	A	A	A	T	T	G	A	A	A	T
D4	A	T	T	G	C	A	A	T	T	G	T	A	G	C
C2	G	C	T	A	C	A	A	T	T	G	C	A	A	T
A	2	1	0	2	1	6	4	0	0	0	3	6	5	0
C	1	2	0	0	5	0	1	0	0	0	1	0	0	3
G	3	0	0	2	0	0	1	0	0	6	0	0	1	0
T	0	3	6	2	0	0	0	6	6	0	2	0	0	3
Consensus:														
	r	y	T	n	c	A	a	T	T	G	n	A	a	Y
	1		3		5		7		9		11	13		(position)

Figure 1A. Relative position of the vir boxes of the *vir* genes: Number at the top indicates the location of the vir boxes relative to the transcription initiation site (+1, according to Das *et al.* (23). Stripped box, required for function; dotted box, partially required for function; dark box, not required; open box, function not known. Number on top of each box identifies individual vir boxes. Brackets in *vir*C, *vir*D and *vir*G indicate regions protected by VirG (13). **Figure 1B.** Sequence comparison of functional vir boxes: Bases that are identical in all six functional boxes are shown in capital letters while those present in four or five of the six are shown in lower case letters. R = purine. Y = pyrimidine. Numbers below the consensus sequence are the residue numbers of the vir box bases. A dyad symmetry is indicated by arrows.

Table 1. Identification of functional vir boxes of *vir*C and *vir*D.

A.

Plasmid	-60	-70	-80	-90	-100	-110	(virD)
	-70	-60	-50	-40	-30	-20	(virC)
pGP264	ACCGAATATAATTGCTACAATTGCAATTTAATTATAACGAATTTGAAATAAAAATAGT						
		C1D5	C2D4	C3D3	C4D2	C5D1	TAAA
pGP265		TCTAGA					
pGP266			TCTAGA				
pGP267				TCTAGA			
pGP294				AAAAGA			
pGP268					TCTAGA		
pGP269						TCTAGA	
pGP270							TCTAGA

B.

Plasmid	ß-galactosidase Activity (U) -AS	+AS
pGP264d	113	13200
pGP265d	113	13000
pGP266.1d	93	105
pGP267d	160	5380
pGP294d	151	13200
pGP268d	96	13600
pGP269d	116	11600
pGP270d.1	168	5990
pGP264c	15	1250
pGP265c	16	1480
pGP266.1c	34	36
pGP267c	30	505
pGP294c	31	1570
pGP268c	20	424
pGP269c	145	1769
pGP270c	24	359

A. The sequence of the vir box region of *vir*C and *vir*D and their mutant derivatives are shown. Numbers at the top indicate location relative to the transcription initiation site of *vir*D and *vir*C. The locations of the vir boxes are indicated below the hexameric core region. The mutations made to each box are shown to the right of the plasmid name that contains the mutation.
B. Plasmids were introduced in *A. tumefaciens* A348. Results reported are averages of three independent assays. Plasmid numbers are followed by either "d" or "c" to indicate orientation of the intergenic region with respect to *lacZ*.

expression indicating that vir box C3D3 is not required for *vir*D expression. The role of the TAAA sequence in *vir*D expression is unclear at present. Mutation of this region (pGP270d) led to a 3 fold reduction in inducible *vir*D expression. How this region, located fairly distant (at position −103) from the promoter sequences, exerts its effect on *vir*D expression is not apparent. DNase I footprinting studies showed that only a small segment of this region, two A residues, were protected by VirG. However, the sequence surrounding it is highly AT rich which is similar to the vir box.

The effect of the same mutations on *vir*C expression was found to be much more complex. Analogous to studies on *vir*D, alteration of the vir box C2D4 (pGP266.1c) abolished *vir*C induction while that of C3D3 (pGP267c) led to attenuated expression as compared to wild type levels (pGP264c). Changing C3D3 to recreate the C2D4 vir box without recreating the C3D3 vir box (pGP294c) restored the inducibility of the *vir*C promoter indicating that vir box C2D4 is required for *vir*C expression and C3D3 is not required for induction *per se*.

Of the other vir boxes destruction of vir box C4D2 (pGP268c) gave an attenuated induction phenotype. This correlates well to a region protected by VirG from DNase I digestion indicating that this box is probably involved in the regulation of *vir*C. However, this box is located at position −45 to −40, just upstream of the −35 region of the *vir*C promoter. Whether the observed affect is due to an effect on the *vir*C promoter cannot be excluded at this time although a promoter mutation is expected to affect both basal and induced level of expression. No effect on the basal level of expression was observed in our study. Alteration of the vir box C5D1 (pGP269c) led to an increase in the basal level of *vir*C expression but had very little effect on the induced level of expression. This box also lies within the −35 region of the *vir*C promoter. Addition of dTCTAGA, a sequence exhibiting some homology to the −35 region of *E. coli* s70 promoters (dTTGACA), to the −35 region could conceivably increase the basal level of expression. Alternately, the increase in basal level may be due to destruction of the binding site for the *ros* repressor which is known to affect transcription of both *vir*C and *vir*D (28). Similar to the studies with *vir*D, alteration of the TAAA sequences (pGP270c) also attenuated induction of *vir*C. This sequence, located just upstream of the −10 region of this promoter, could conceivably affect the promoter but we have not distinguished between disruption of the promoter and disruption of VirG binding sites.

Vir box B2 contributes to *vir*B induction but is not sufficient.

In a previous study we demonstrated that vir box B1 is an absolute requirement for *vir*B expression (15). This study, however, could not elucidate the role of vir box B2 in *vir*B expression. Site directed mutagenesis was used to alter the central four bases of vir boxes B1 and B2 and the effects of these alterations on *vir*B expression were monitored using a *vir*B:*lacZ* translational fusion (Table 2). As expected, mutation of the vir box B1 (pGP245) led to complete loss of *vir*B expression. The same change to vir box B2 did not have nearly as drastic an effect (pGP244.7). This promoter was inducible but only to 40% of the level of the wild type promoter (pAD1092) indicating that it is only partially required for induction.

The vir box consensus palindrome dGTTTCAATTGAAAC is sufficient to confer inducibility to a deleted *vir*E promoter

The *vir*E promoter contains a single vir box that is required for *vir*E expression and is protected by VirG in an *in vitro* DNase I protection assay (12). To determine what constitutes a vir box we chose to use this promoter so that the analysis would not be complicated by additional vir boxes. A *vir*E:*lacZ* translational fusion was constructed to monitor *vir*E expression. The vir box sequence was altered by site-specific and insertional mutagenesis. The effects of these mutations on *vir*E expression is summarized in Table 3.

Agrobacterium harboring plasmid pDS1c that contains −500 to +425 of the *vir*E sequence showed inducible expression of *vir*E (line 1) indicating that sequences required for *vir*E expression are encoded within this region. This strain exhibited an unusually high basal level of expression compared to that of other *vir* promoters. A similiar high basal level of *vir*E expression was observed from the Ri plasmid A4 *vir*E operon (18). Both mutation of the central six bases of the vir box (pGP275) and deletion upto residues −47 (pΔ27c) led to the complete loss of inducible *vir*E expression confirming earlier observations of Jin *et al.* (12) that the vir box E1 is required for *vir*E expression. The *vir*E vir box does not exactly match the consensus vir box palindrome. To determine how the consensus palindrome would functionally compare to the native sequence plasmid pGP322 was constructed. An alteration of two divergent bases of pDS1c (positions 4 and 7 of the vir box, see Figure 1B for numbering) led to a two fold increase in the level of induction. These alterations had very little effect on the uninduced level of expression indicating that the observed effect was due to an improvement in the vir box. In another mutant a G to T change at position 10 of the vir box (pGP323) led to a four fold reduction in induction indicating that a G residue at this position is preferred for maximal induction.

Our studies with the *vir*C and *vir*D promoters indicated that at least three additional residues downstream from the conserved hexameric core region are necessary for a functional vir box. When these bases were altered a significant reduction in *vir*C and *vir*D expression was observed (Table 1, plasmids pGP267c,d). This suggests that the vir box is at least dodecameric in size. To determine if the dodecameric region constitutes a vir box we introduced an adaptor DNA oligomer containing the dodecameric consensus palindrome into pΔ27c (pGP304). This led to a five fold increase in *vir*E expression in the presence of AS. This level, however, is lower than that observed in strains containing pGP322. The DNA sequence of the vir box region of pGP322 and pGP304 differ in regions upstream and downstream from the vir box. Our deletion mutagenesis studies indicated that sequence upstream of -66 is not required for *vir*E induction (unpublished results). We therefore speculated that the changes downstream from the vir box region may contribute to the low level of induction in strains harboring pGP304. To test this possibility we constructed pGP321 which has the G residue at position 14 changed to the consensus C residue. Comparison of DNA sequences of the various required vir box sequences indicated that this position is usually a pyrimidine residue. This alteration led to an additional 2 fold increase in induction indicating that a C residue at position 14 is preferred for maximal *vir*E expression. A comparison of DNA sequence of the vir box region in pGP321 and pGP322 show that sequences downstream of residue 14 and upstream of residue 2 are different. Since strains containing either of these plasmids exhibit similiar levels of *vir*E expression the vir box must be contained within these 13 residues.

Table 2. Role of vir box B2 in *vir*B expression

Plasmid	Relevant DNA sequence -70 -60 -50 -40 -30	β-galactosidase Activity (U) −AS	+AS
pAD1092	CGCTTCAATTGAAATCATAAAGAAGCAATTGAAAATTTTCG	11	1770
pGP245	CCGG	10	11
pGP244.7	CCGG	9	729
vir box:	B1 B2		

All plasmids were introduced into *A. tumefaciens* A348. Results reported are averages of three independent assays. Relevant sequence around the vir boxes are numbered with respect to the start of *vir*B transcription. Only the altered bases are shown for pGP245 and pGP244.7.

Table 3. Reconstruction of a functional vir box.

Plasmid	Relevant DNA sequence -70 -60 -50 -40	Deletion Endpoint	β-galactosidase Activity (U) −AS	+AS
pDS1c	.....GCCTGCTCATTGCAGTTGAAACGCGATATCCGTTTCA	−500	1130	14000
pGP275	.....GCCTGCTCATTGTCTAGAAACGCGATATCCGTTTCA		495	447
pGP322	.....GCCTGCTCATTTCAATTGAAACGCGATATCCGTTTCA		767	25200
pGP323	.....GCCTGCTCATTTCAATTTAAACGCGATATCCGTTTCA		559	5800
pΔ27c	vector.....GAATTCCCGTTTCA	-47	1100	1100
pGP304	GAATTAGCTTTCAATTGAAAGCTAATTCCCGTTTCA		1910	11100
pGP321	GAATTAGGTTTCAATTGAAACCTAATTCCCGTTTCA		1260	25600

All plasmids were introduced into *A. tumefaciens* A348. Results are averages of two independent assays. Relevant sequence of the *vir*E operon is numbered with respect to the start of *vir*E transcription. Deletion endpoint refers to the amount of *vir*E sequence upstream of the start of *vir*E transcription. Changes from wild type sequence are underlined.

DISCUSSION

A vir box sequence is essential for *vir* gene expression (15). Several *vir* genes, *vir*B, *vir*C, *vir*D and *vir*G contain multiple vir boxes. In this study we examined the role of individual vir boxes in *vir* gene induction by acetosyringone. *Vir*B and *vir*C were found to utilize two vir boxes while *vir*D used only one. *Vir*B and *vir*C each had one major vir box that was absolutely required for induction and had a second that contributed to induction but was not absolutely required. The major vir box of *vir*C was the same vir box used by the divergent *vir*D operon. This is possible because the vir box sequence has a dyad symmetry. Symmetry of binding site sequence indicates that the binding protein is probably dimeric or tetrameric (29). Thus a dimer or tetramer of VirG could bind at this common box and be positioned to affect transcription in both directions.

We found a very good correlation between the vir boxes required for induction and those protected by VirG in a DNase I protection assay. We reported earlier that two vir boxes were protected between the *vir*C and *vir*D operons (13). In the present study we show that these two boxes are required for induction of the respective genes while those not protected by VirG are not functionally required. Winans (17) has observed that two of the three vir boxes of *vir*G are required for its induction by AS. We have reported earlier that the same two vir boxes are bound by VirG *in vitro* (13).

A comparison of the DNA sequences of the functional vir boxes yields a consensus sequence is 5'd ryTncAa TTGnAaY (bases present in all six required vir boxes are capitalized while those present in 4 or 5 out the 6 boxes are shown in lower case letters; R = A or G; Y = C or T). A homolog of this sequence restores inducibility of a non-inducible *vir*E promoter (contains a deletion of the native vir box) in a reconstruction experiment (Table 3). These results support the hypothesis that the consensus vir box sequence is functionally active. The consensus sequence has a dyad symmetry (arrows, Figure 1B) which allows the vir box to be split into two half-sites. Of these the downstream half-site is much more conserved than the upstream half-site. The vir box

consensus for the Ri plasmid RiA4 *vir* genes was reported to be 5′ dTG(A/T)AA(C/T) (18) which is very similar to the Ti plasmid vir box half-site. It has been proposed that this sequence is sufficient for VirG induction of the Ri *vir* genes (18,30). Our previous analysis of the *vir*B promoter would indicate that a half-site is not sufficient for induction of the Ti plasmid *vir*B operon. We found that inducible expression of *vir*B was lost in a mutant derivative (Δ100, Figure 1, ref. 15) which had lost the upstream half-site of vir box B1 but had no alterations in the downstream half-site.

The functional vir boxes of a majority of the virulence genes have been identified (12,15,17, this study). These vir boxes are located throughout the upstream regions from −20 to −80. For a vir box to be functional it should have a primary sequence compatible with the VirG binding site sequence and be placed in a proper position for interaction with RNA polymerase. The spacing of vir boxes has been largely unaddressed. The only work on this problem was with the *vir*B operon and it was complicated by the presence of vir box B2 (15). In the closely related OmpR/EnvZ system of *E. coli* there are multiple OmpR binding sites upstream of the *omp*C promoter. However, one properly located binding site is sufficient (31). Displacement of the OmpR binding sites is allowed if the displacements are at intervals of 10 or 11 bases indicating that the face of the helix that OmpR binds to is important for function (32). The same may be true in the *Agrobacterium vir* system. The functional vir boxes of all the *vir* operons (except *vir*G) tend to lie at approximately 10 base intervals from each other (Figure 1A). This would indicate that they all lie on the same face of the DNA helix and could all interact with RNA polymerase in a similiar manner. *Vir*G is an exception as it does not follow this interval spacing and has one of its functional vir boxes located between the −10 and −35 regions. This suggests that the contacts between VirG and RNA polymerase may be different in the *vir*G promoter. In addition, in *vir*B, *vir*C and *vir*G , where a second vir box is required, the spacing between the two vir boxes is 20, 11 and 33 base pairs, respectively. This would suggest that the two vir boxes lie on the same face of the DNA helix allowing interaction between the bound VirG molecules at the two sites.

ACKNOWLEDGEMENTS

We thank Ann Vogel and John Larkin for a critical reading of the manuscript and Dominique Shimizu for constructing pDS1c. This study was supported by a grant from the National Institute of Health (GM 37555), a Junior Faculty Research Award from the American Cancer Society (JFRA 170) and a McKnight Land-Grant Professorship Award from the University of Minnesota. GP is a Graduate School Dissertation Fellow of the University of Minnesota.

REFERENCES

1. Stachel, S., and Zambryski, P. (1986) Cell, 46, 325−333.
2. Rogowsky, P., Close, T., Chimera, J., Shaw, J., and Kado, C. (1987) J. Bacteriol., 169, 5101−5112.
3. Winans, S., Kerstetter, R., and Nester, E. (1988) J. Bacteriol., 170, 4047−4054.
4. Winans, S., Ebert, P., Stachel, S., Gordon, M., and Nester, E. (1986) Proc. Natl. Acad. Sci. USA, 83, 8278−8282.
5. Leroux, B., Yanofsky, M., Winans, S., Ward, J., Ziegler, S., and Nester, E. (1987) EMBO J., 6, 849−856.
6. Ronson, C., Nixon, B., and Ausubel, F. (1987) Cell, 49, 579−581.
7. Stock, J., Ninfa, A., and Stock, A. (1989) Microbiol. Rev., 53, 450−490.
8. Jin, S., Roitsch, T., Ankenbauer, R., Gordon, M. and Nester, E. (1990) J. Bacteriol., 172, 525−530.
9. Huang, Y., Morel, P., Powell, B. and Kado, C. (1990) J. Bacteriol., 172, 1142−1144.
10. Stachel, S., Messens, E., Van Montagu, M., and Zambryski, P. (1985) Nature (London), 318, 624−629.
11. Jin, S., Prusti, R., Roitsch, T., Ankenbauer, R. and Nester, E. (1990) J. Bacteriol., 172, 4945−4950.
12. Jin, S., Roitsch, T., Christie, P., and Nester, E. (1990) J. Bacteriol., 172, 531−537.
13. Pazour, G., and Das, A. (1990) J. Bacteriol., 172, 1241−1249.
14. Winans, S., Jin, S., Komari, T., Johnson, K., and Nester, E. (1987) in Plant Molecular Biology, von Wettstein, D and Chua, N.-H., eds. Plenum Press, NY, pp. 573−582.
15. Das, A., and Pazour, G. (1989) Nucleic Acids Res., 17, 4541−4550.
16. Steck, T., Morel, P. and Kado, C. (1988) Nucleic Acids Res., 16, 8736.
17. Winans, S. (1990) J. Bacteriol., 172, 2433−2438.
18. Aoyama, T., Takanami, M., and Oka, A. (1989) Nucleic Acids Res., 17, 8711−8725.
19. Vieira, J., and Messing, J. (1987) Methods in Enzymology, 153, 3−11.
20. Jayaswal, R., Veluthambi, K., Gelvin, S., and Slightom, J. (1987) J. Bacteriol., 169, 5035−5045.
21. Kunkel, T. (1985) Proc. Natl. Acad. Sci. USA, 82, 488−492.
22. Sanger, F., Nicklen, S., and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463−5467.
23. Das, A., Stachel, S., Allenza, P., and Nester, E. (1986) Nucleic Acids Res., 14, 1355−1364.
24. Winans, S., Allenza, P., Stachel, S., McBride, K., and Nester, E. (1987) Nucleic Acids Res., 15, 825−836.
25. Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. (1980) Proc. Natl. Acad. Sci. USA, 77, 7347−7351.
26. Chilton, M.-D., Currier, T., Farrand, S., Bendich, A., Gordon, M. and Nester, E. (1974) Proc. Natl. Acad. Sci. USA, 71, 3672−3676.
27. Miller, J. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
28. Close, T., Rogowsky, R., Kado, C., Winans, S., Yanofsky, M., and Nester, E. (1987) J. Bacteriol., 169, 5113−5118.
29. Schleif, R. (1988) Science, 241, 1182−1187.
30. Aoyama, T., and Oka, A. (1990) FEBS Letters, 263, 1−4.
31. Maeda, S., and Mizuno, T. (1990) J. Bacteriol., 172, 501−503.
32. Maeda, S., Ozawa, Y., Mizuno, T., and Mizushima, S. (1988) J. Mol. Biol., 202, 433−441.