Volume 14 Number 3 1986 Nucleic Acids Research

Promoters of *Agrobacterium tumefaciens* Ti-plasmid virulence genes

A.Das[1], S.Stachel[2], P.Ebert[3], P.Allenza[4], A.Montoya[5] and E.Nester

Department of Microbiology and Immunology SC-42, University of Washington, Seattle, WA 98195, USA

Received 26 August 1985; Revised 18 November 1985; Accepted 20 December 1985

ABSTRACT

The DNA sequences of the promoter and 5' upstream regions of six Agrobacterium tumefaciens Ti-plasmid encoded virulence (vir) genes were determined. The transcription initiation sites were mapped by the S_1 nuclease protection assay. In the -10 region, the vir promoters share a consensus sequence that is homologous to a DNA sequence found in the same region of E. coli promoters. In contrast, the -35 region sequences are variable. Several vir genes contain two common hexanucleotide sequences, 5'CGAGTA3' and 5'GCAATT3'. Translation initiation codons for all vir genes, except virG, are preceded by sequences homologous to the ribosome binding site sequences found in E. coli.

INTRODUCTION

The soil bacterium Agrobacterium tumefaciens incites tumors on most dicotyledonous plants by transferring a segment of plasmid-borne DNA into plant genome (reviewed in ref. 1,2). Expression of genes encoded in the transferred DNA (T-DNA) following its integration into the plant nuclear genome leads to the tumorous phenotype. The DNA transfer functions lie outside the T-DNA in an adjacent segment, termed the virulence (vir) region (3-5). This region, encompassing about 35 Kb of DNA, contains at least six complementation groups: virA, virB, virC, virD, virE and virG (5, S.S. and E.N., manuscript in preparation). Genetic studies using a Tn3 - lacZ transposon have demonstrated that the expression of several vir genes of an octopine type Ti plasmid, pTiA6, are inducible upon cultivation with plant cells (6). Two vir operons, virA and virG, are expressed in Agrobacterium grown in culture. Appreciable expression of virB, virC, virD and virE was observed only upon cocultivation of Agrobacterium with plant cells. Such induction also required the presence of both virA and virG gene products (7, S.S. and E.N., manuscript in preparation). Inducible expression of the virC locus (homologous to virD of Klee et al. (4)) of a nopaline-type Ti plasmid has also been reported (8).

To better understand the molecular basis of expression and regulation of the Agrobacterium virulence genes we analyzed the DNA sequences upstream of the transcription initiation sites. The promoter region sequences show considerable homology to the consensus sequences for E. coli promoters. All vir genes were found to contain one conserved hexanucleotide sequence while virB, virC, virD and virE contained a second conserved hexanucleotide sequence.

MATERIALS AND METHODS

Induction of Agrobacterium, RNA isolation and S_1 mapping:

Agrobacterium strain A348 containing plasmid pTiA6 was grown in liquid mannitol-glutamate medium at 30°C (9). 5 ml of a culture grown overnight was centrifuged, resuspended in 0.5 ml of the same medium and added to 50 ml of a 3 day old subculture of Nicotiana tabacum suspension culture (cell line NT1). The plant cell-bacterial mixture was incubated in a rotary shaker for 4-6 hrs. and plant cells were then separated from bacteria by 2 rounds of low- speed centrifugation at 600 rpm for 2-3 mins. each. Bacterial cells were harvested by centrifugation at 5000 rpm for 5 mins. at 4°C in a Sorvall SS-34 rotor. Total RNA was isolated from induced and uninduced Agrobacterium by the hot phenol method of Aiba et. al. (10). S_1 nuclease mapping was essentially according to published procedures (11-13). Briefly, 50-100 µg total RNA and 20-30,000 cpm of 5'-[p^{32}] labelled double-stranded DNA probe (sp. ac. 0.5-1x10^7 cpm/µg) were used in each experiment. RNA and probe DNA were co-precipitated with ethyl alcohol, dried and resuspended in 30 µl hybridization buffer (80% formamide, 0.4M NaCl, 40mM phosphate buffer, pH4.6, 1mM EDTA), denatured at 80°C for 10-15 mins. and hybridized overnight at 48°C. Following hybridization 300 ul S_1 nuclease buffer (13) containing 500-1,000 u/ml of S_1 nuclease was added and incubation was continued at 37°C for 1 hr. S_1 resistant nucleic acids were isolated by phenol extraction followed by ethanol precipitation. Samples were analyzed in denaturing acrylamide (6%) sequencing gels next to molecular weight marker DNAs and a sequencing ladder.

Other procedures:

Molecular cloning, plasmid preparation and other standard procedures were according to Maniatis et. al. (13). DNA sequencing was performed by the chemical method of Maxam and Gilbert (14) or by the dideoxy chain termination method of Sanger and co-workers (15). Restriction endonucle-

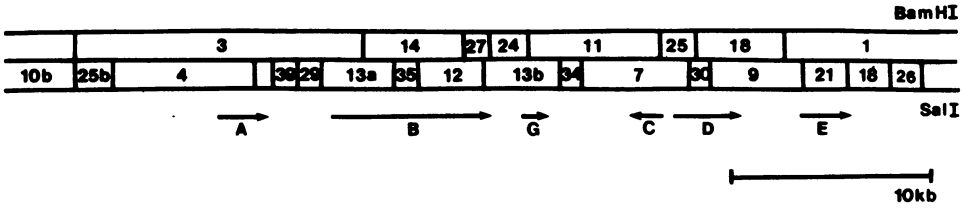

Fig. 1. Physical Map of the Virulence Region of pTiA6. The BamH1 and Sal1 restriction endonuclease maps of the octopine Ti plasmid virulence region are shown. The letter designations are for individual vir genes and the arrows indicate the direction of transcription. The T-DNAs lie to the right of the BamH1 restriction fragment 1.

ases and other enzymes were from Bethesda Research Laboratories, New England Biolabs and International Biotechnologies. Radionucleotides were from New England Nuclear.

RESULTS

Extensive transposon-mediated mutagenesis using the transposons Tn5, Tn3 and Tn3 - lacZ has defined the location, boundary, and direction of transcription of the genes of the virulence region of Agrobacterium Ti plasmid pTiA6 (Fig. 1) (3-5, S.S. and E.N., manuscript in preparation). DNA fragments encompassing the 5' regions of the vir genes were subcloned into pUC and m13 vectors (16). The DNA sequences of the putative promoter regions were then determined. The sequence information was used to delineate restriction fragments that were subsequently used as probes in S_1 nuclease protection experiments to define the 5' termini that presumably are the transcription initiation sites (Table 1, Fig. 2).

virA, virC, and virD genes encoded mRNAs with unique transcription initiation sites (Fig. 2, lanes 3,6,7) while virB specified two transcripts in approximately equimolar amounts that started 1 base pair apart (Fig. 2, lane 14). virE and virG mRNAs showed multiple bands in S_1 protection studies (Fig. 2, lanes 11,16). Further analysis of virE mRNA start sites using different experimental conditions, e.g. varying amounts of RNA, varying concentrations of S_1 nuclease and different S_1 digestion temperatures suggested that the largest protected fragment corresponded to the major transcription initiation site (data not shown). The molar ratios of the other fragments were highly variable under different experimental conditions suggesting that multiple bands resulted from protection of partially degraded RNAs. However, secondary initiation at

Table 1. Probes for S_1 Nuclease Mapping.

Locus	Probe length, bp	Region	Restriction fragment	Putative initiator codon at position
virA	530	-357 to +182	*DdeI - HindIII	+ 38
virB	367	-106 to +260	*HpaII - AvaII	+ 64
virC	540	-288 to +251	*BamH1 - XmaIII	+ 81
virD	540	-378 to +161	*XmaIII - BamH1	+ 61
virE	555	-195 to +359	*Bst N1 - Sal1	+ 396
	797+	-195 to +601	*DdeI - Sal1	+ 396
virG	452	-145 to +306	*XhoI - BglII	+ 48

* Indicates the 5' end that was radio-labelled by T4 polynucleotide kinase. The transcription initiation site has been numbered +1.
+ Data not shown.

these sites or involvement of RNA processing cannot be ruled out unequivocally. The smaller band in virG studies (Fig. 2, lane 11) may likewise be artifactual because this site mapped past the translation initiation codon of the largest open reading frame (ORF) (Ebert et. al., submitted).

The DNA sequences, extending from 140 base pair (bp) preceding to 10 bp past the transcription initiation sites, of the six vir promoters are shown in Fig. 3. In the non-transcribed region, all genes shared common sequences in their -10 region but lacked similarity in the sequences around the -35 region (underlined sequences, Fig. 3). The consensus -10 region sequence of the Agrobacterium vir genes is 5'T/GNTAAT/C3', a sequence homologous to consensus -10 region sequences, 5'TATAAT3', found in E. coli promoters (17).

The DNA sequence of the upstream region of each vir gene was analyzed for other common features. Two hexameric sequences were present in several genes. One, 5'CGAGTA3' (hereafter written as CGAGTA), was found at positions -31, -24 and -34 (with respect to the C residue) of virB, virD, and virE, respectively (overlined sequences, Fig. 3). virC had a similar sequence, CGAATA, at position -70. virG and virA lacked any sequence homologous to this in the sequenced regions. A second hexameric sequence,

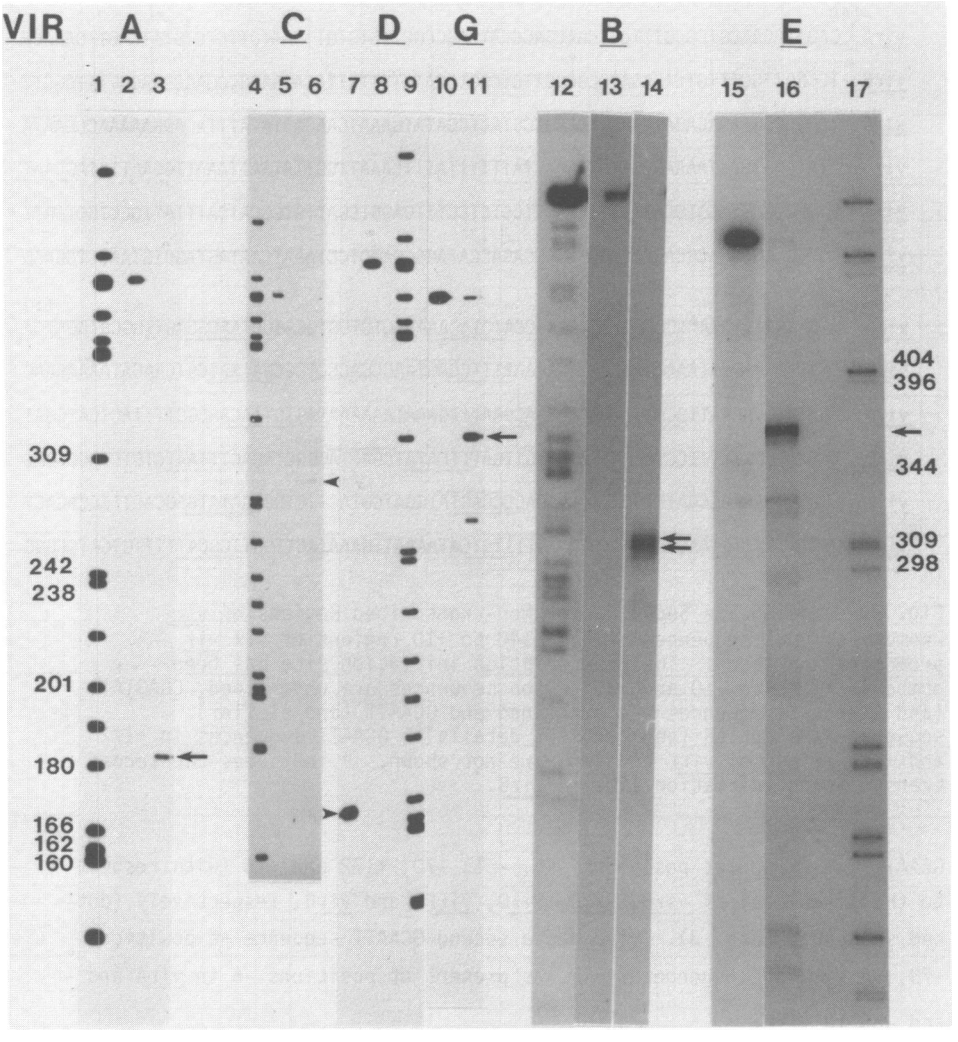

Fig. 2. S_1 Nuclease Mapping of *vir* mRNAs. Experimental details were as described in Materials and Methods. Lanes 1,4,9 and 17 are molecular weight markers; lane 12 is a G reaction of chemical sequencing (14) with the *virB* probe; lanes 2,5,8,10,13 and 15 are probes without RNA and without S_1 nuclease; lanes 3,6,7,11,14 and 16 are probes plus RNA from induced bacteria plus S_1 nuclease. Control reactions with wheat germ tRNAs and data with RNAs from uninduced bacteria are not shown. Except for *virB* all others have an estimated error of ± 2 bp. The marker DNAs in lanes 1, 4 and 9, and in lane 17, were double digests of pBR322 plasmid DNA with restriction endonucleases HpaII, DdeI and HpaII, HinfI, respectively.

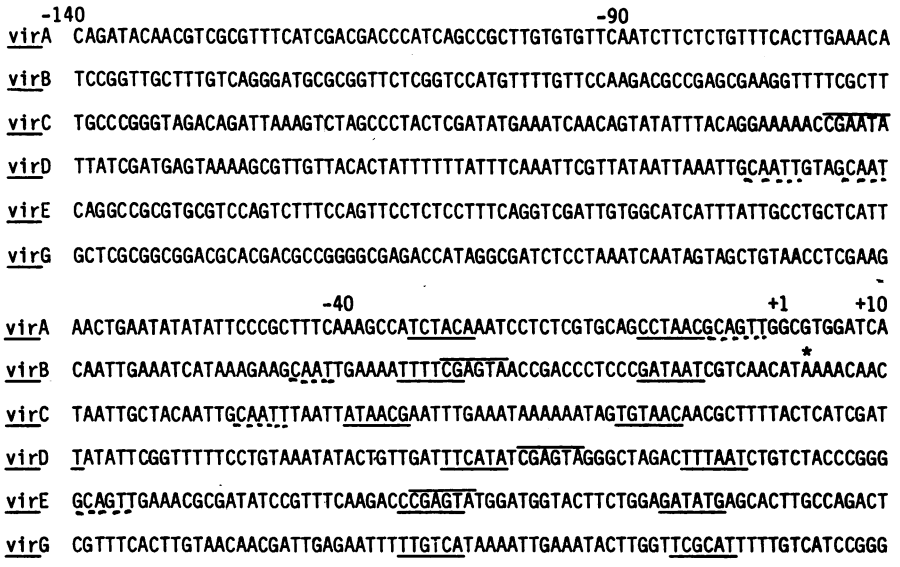

Fig. 3. Nucleotide Sequences of Non-transcribed Regions of vir Promoters. DNA sequences of the -140 to +10 regions of six vir promoters are shown. The transcription initiation site has been numbered +1. The -10 and -35 region sequences are underlined, CGAGTA (and similar) sequences are overlined and GCAATT (and similar) sequences are dotted (see text for details). GCAATT sequences in +172 and +15 regions of virE and virG are not shown. * indicates the second transcription initiation site in virB.

GCAATT, was found at positions -46, -51, -70, +172 and +15 (with respect to the G residue) of virB, virC, virD, virE, and virG, respectively (dotted sequences, Fig. 3). virD had a second GCAATT sequence at position -79. A similar sequence GCA$\frac{G}{C}$TT was present at positions -6 in virA and -4 and -65 in virE.

The DNA sequences of several hundred base pairs 3' to the transcription initiation sites were determined to establish the N-terminal segments of encoded polypeptides. Analysis of the sequences preceding the putative initiator ATG triplet showed that all ORFs, except virG, encoded polypeptides that are preceded by a sequence highly homologous to the ribosome binding site sequences found in E. coli (Table 2).

DISCUSSION

Ti-plasmid mediated DNA transfer into plant cells is dependent on the expression of the vir genes. These genes are expressed in the bacterium

Table 2. Ribosome Binding Site Sequences of *vir* Genes.

Locus	Sequence (-20 to -10)	Initiator codon
virA	GGCACGAGGA AGTTAGTGCG	ATG
$virB_1$	GCGAGCTAAG GAGATAAGGT	ATG
$virB_2$	TACTAAGGAG GTCCGCATTG	ATG
$virC_1$	AGGACAGGCG GGATCAGAAT	ATG
$virC_2$	AAAATCTTGG AGGCTTGAAG	ATG
virD	CCGCTTTTTT GGAGGAAGCT	ATG
virE	TGGTCAGAAG GAGTGAGACG	ATG
virG	AACTGCCCAT TTAGCTGGAG	ATG
E. coli	TAAGGAGGTG..5-9 bp...	ATG

Sequences homologous to the *E. coli* Shine-Dalgarno sequence (26) have been underlined. The subscripts indicate the different ORFs in each *vir* locus. The numbers increase from 5' to 3' direction.

but most of them require a plant factor(s) for their expression (6). Of all *vir* genes, only *virA* shows constitutive expression. *virG* also is expressed in the bacteria but it appears to be positively auto-regulated. *virB*, *virC*, *virD* and *virE* show inducible expression only in the presence of *virA*, *virG* and plant factor(s) (S.S. and E.N., manuscript in preparation).

The DNA sequence analyses of 5' ends show that the promoter regions of all *vir* promoters share sequence homology to one another in the -10 region and less so in the -35 region. The -10 region consensus sequence of the vir promoters, $\frac{T}{G}NTAA\frac{T}{C}$, is similar to TATAAT, the sequence found in the same region of many *E. coli* and bacteriophage promoters (reviewed in ref. 17). However, the two consensus sequences exhibit significant differences as well. A major difference is that while the last T residue of TATAAT sequence is invarient in 46 different *E. coli* and coliphage promoters, the residue in the same position is not a T residue in 3 of the 6 *vir* promoters (Fig. 3). A second is that the most conserved residues in the

hexameric sequence in E. coli promoters are the T, A and T residues at positions 1, 2 and 6 (relative to the 5' most T residues), respectively. In contrast, these three residues are the least conserved in the Agrobacterium promoters.

The sequences in the -35 region exhibit a varying degree of homology to one other as well as to other bacterial promoters. Two promoters, virA and virG, show the most homology with the consensus bacterial -35 region sequence, TTGACA. Interestingly, these two genes are the only vir genes that are expressed in bacteria grown in culture. The promoters of virB, virC, and virE show little homology to the consensus -35 region sequence. In contrast to virA and virG, these genes are not expressed in uninduced bacteria at any detectable levels. Thus there appears to be a correlation between efficient transcription in Agrobacterium and the consensus sequences in the -35 and -10 regions. The sixth vir promoter, virD, shows some variation from the above pattern. Although both -10 and -35 region sequences of virD show significant homology to the consensus sequences of bacterial promoters, this gene is not expressed in Agrobacterium to any appreciable extent. One possible explanation is that the -35 region sequence is sufficiently different from the consensus sequence of constitutive Agrobacterium promoters to result in highly reduced promoter activity.

The non-transcribed region of several vir genes contains two conserved hexameric sequences. One, CGAGTA (or a homologous sequence), is present in the non-transcribed regions of virB, virC, virD and virE genes (Fig. 3). Coincidentally, each of these promoters is strictly inducible. No similar sequence is present in the sequenced regions of the constitutive virA and virG genes. The second, GCAATT, is present in the non-transcribed regions of virB, virC and virD and in the transcribed region of virE and virG (Fig. 3). No such sequence is found in virA. However, highly homologous sequences are found in non-transcribed regions of virA and virE genes. Among the vir promoters the CGAGTA sequence is located between -70 and -20, and the GCAATT sequence is present between -80 and +180, with respect to the transcription initiation sites. Whether these sequences have any regulatory or other significance is not clear at present. Short sequences located upstream of the transcription initiation sites have been implicated in positive regulation of the galactose, lactose and other operons by the cyclic-AMP binding protein (CAP), arabinose operon by the araC protein, and bacteriophage lambda P_L and P_R promoters by the cI and cII proteins, respectively (18-23). In CAP regulated promo-

ters the position of the regulatory segments is variable as well (20). Deletion mutation studies will determine if any of the hexanucleotide sequences found conserved in the vir promoters has a similar role in the positive regulation of the vir genes.

The putative translation initiation codon of all vir genes, sequenced in our laboratory and elsewhere (24,25), is preceeded by sequences similar to the Shine-Dalgarno sequence found in E. coli (Table 2) (26). This suggests that the 3' end of 16S RNAs of Agrobacterium and E. coli share considerable sequence homology. One gene, virG, totally lacks a ribosome binding site sequence. Some regulatory proteins in E. coli and bacteriophage λ also lack these sequences (27,28). It has been suggested that lack of a ribosome binding site sequence serves as a translational regulatory mechanism to prevent overexpression of gene products that are required in small quantities (29). All vir genes, except virE, contain a short, 30-80 bp, non-translated region (Table 1). virE, on the other hand, has a 400 bp non-translated region. The reason for such a large non-translated region in the latter is not apparent.

ACKNOWLEDGEMENTS

We thank Lisa Albright, Frank White and our other colleagues for a critical reading of the manuscript, Marty Yanofsky for providing us with unpublished DNA sequences and Ramona Finical for excellent typing of the manuscript. This work was supported by grants from the National Institutes of Health (GM-32618-14) and U.S. Department of Agriculture (84-CRCR-1-1438). P.A. was a NIH-postdoctoral trainee (GM09050-02).

Present addresses: [1]Department of Biochemistry, St. Paul, MN 55108, [2]Laboratorium Genetika, Rijksuniversiteit Gent, Gent, Belgium, [3]Institute of Biological Chemistry, Washington State University, Pullman, WA 99164, [4]Signal Research Center, Des Plaines, IL 60016 and [5]Ciba-Geigy, Research Triangle Park, NC 27709, USA

REFERENCES

1. Nester, E., Gordon, M., Amasino, R. and Yanofsky, M. (1984) Ann. Rev. Plant Physiol. 35, 387-413.
2. Hooykas, P. and Schilperoort, R. (1983) Adv. Genetics 22, 209-283.
3. Garfinkel, D. and Nester, E. (1980) J. Bacteriol. 144, 732-743.
4. Klee, H., Gordon, M. and Nester, E. (1982) J. Bacteriol. 150, 327-331.
5. Klee, H., White, F., Iyer, V., Gordon, M. and Nester, E. (1983) J. Bacteriol. 153, 878-883.
6. Stachel, S., An, G., Flores, C. and Nester, E. (1985) EMBO J. 4, 891-898.
7. Stachel, S., An, G. and Nester, E. (1984) J. Cellular Biochem. Suppl. 8B, pp. 64.

8. Okker, R., Spaink, H., Hiue, J., van Brussel, T., Lugtenberg, B. and Schilperoort, R. (1984) Nature 312, 564-566.
9. Chilton, M.-D., Currier, T., Farrand, S., Bendich, A., Gordon, M. and Nester, E. (1974) Proc. Natl. Acad. Sci. USA 71, 3672-3676.
10. Aiba, H., Adhya, S. and de Crombrugghe, B. (1981) J. Biol. Chem. 256, 11905-11910.
11. Berk, A. and Sharp, P. (1977) Cell 12, 721-732.
12. Favaloro, J., Treisman, R. and Kamen, R. (1980) Methods Enzymol. 65, 719-749.
13. Maniatis, T., Fritsch, E. and Sambrook, J. (1982) Molecular cloning: A Laboratory Manual, Cold Spring Harbor Lab, NY.
14. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
15. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Messing, J. and Viera, J. (1982) Gene 19, 269-276.
17. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
18. Gilbert, W. (1976) in RNA Polymerase, R. Losick and M. Chamberlin, eds. pp 719-730. Cold Spring Harbor Lab., NY.
19. de Crombrugghe, B. and Pastan, I. (1978) in The Operon, J. Miller and W. Reznikoff, eds. pp 304-324. Cold Spring Harbor Lab., NY.
20. de Crombrugghe, B., Busby, S. and Buc, H. (1984) Science 224, 831-838.
21. Greenfield, L., Boone, T. and Wilcox, G. (1978) Proc. Natl. Acad. Sci. USA 75, 4724-4728.
22. Ptashne, M. (1978) in The Operon, J. Miller and W. Reznikoff, eds. pp 325-343. Cold Spring Harbor Lab., NY.
23. Rosenberg, M., Court, D., Shimalake, H., Brady, C. and Wulff, D. (1978) in The Operon, J. Miller and W. Reznikoff, eds. pp 345-371. Cold Spring Harbor Lab., NY.
24. Hille, J. (1983) Ph.D. thesis, University of Leiden, Leiden, The Netherlands.
25. Hagiya, M., Close, T., Tait, R. and Kado, C. (1985) Proc. Natl. Acad. Sci. USA 82, 2669-2673.
26. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342-1346.
27. Gunsalus, R. and Yanofsky, C. (1980) Proc. Natl. Acad. Sci. USA 77, 7117-7121.
28. Humayun, Z., Zeffrey, A. and Ptashne, M. (1977) J. Mol. Biol. 112, 265-277.
29. Kelley, R. and Yanofsky, C. (1982) Proc. Natl. Acad. Sci. USA 79, 3120-3124.