Volume 13 Number 15 1985 Nucleic Acids Research

Nucleotide sequence and mutational analysis of an immunity repressor gene from *Bacillus subtilis* temperate phage φ105

Patrick Dhaese, Jozef Seurinck[1], Bea De Smet and Marc Van Montagu*

Laboratorium voor Genetica, Rijksuniversiteit Gent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

Received 12 June 1985; Accepted 19 July 1985

ABSTRACT

We have identified and sequenced a bacteriophage φ105 gene encoding an immunity repressor, the first to be characterized from a temperate phage infecting a Gram-positive host. Using superinfection immunity as an assay for repressor function, the φ105 repressor gene was located within a 740-bp PvuII-HindIII subfragment near the left end of the φ105 EcoRI-F fragment. We show that the repressor is specified by the 5'-proximal coding sequence of a translationally overlapping gene pair, transcribed from right to left on the conventional φ105 map. Comparison of its amino acid sequence (146 residues) with that of a large number of Gram-negative bacterial and phage repressors revealed a putative DNA-binding region between positions 20 and 39. The coding region is preceded by a strong Shine-Dalgarno sequence 5' AAAGGAG 3'. Deletion analysis of the 5'-flanking DNA allowed to identify transcriptional control elements. Their structure, 5' TTGTAT 3' at -35 and 5' TATAAT 3' at -10, strongly suggests that the φ105 repressor gene is transcribed by the major vegetative form of B. subtilis RNA polymerase, as would be expected for an early phage gene.

INTRODUCTION

Through the impressive amount of work on the lambdoid coliphages [reviewed in 1], it is well documented that the lysogenic response to infection by a temperate phage requires the synthesis of a phage-encoded repressor. This regulatory protein binds to specific DNA sequences, termed operators, thereby blocking the expression of lytic genes. As such, the repressor is also responsible for the immunity of a lysogen to superinfection by the same or a closely related temperate phage.

The Bacillus subtilis phage φ105 (genome size, 38 kb) most probably uses a similar pathway for the establishment of lysogeny, but until recently only little was known on the genetic organization of its immunity region(s) [2,3]. Combined evidence obtained from the mapping of clear-plaque deletion mutants [4] and molecular cloning [5,6] suggested that a functional φ105 repressor gene is located within the 3.2-kb EcoRI-F fragment (Figure 1). Recently, we have assigned this repressor gene more precisely to a

1.1-kb EcoRI-HindIII subfragment at the left extremity of EcoRI-F [7]. We also identified a repressor-controlled φ105 promoter, which was mapped to a 662-bp BclI fragment, partially overlapping the repressor fragment at its right end [3,7]. Both these elements, repressor and early promoter, were used in the construction of a thermo-inducible gene expression system for B. subtilis [7].

Here, we report a detailed analysis, by molecular cloning, DNA sequencing and deletion mutagenesis, of this φ105 repressor gene and its flanking regions. We show that the repressor forms part of a pair of overlapping genes. In addition the amino acid sequence of the φ105 repressor polypeptide was compared to that of other repressors in order to identify regions possibly involved in DNA-binding.

MATERIALS AND METHODS

Bacterial strains, phage, and plasmids

The following B. subtilis strains were used in this work : B. subtilis 168 (trpC2) and its φ105 wild-type lysogen, and BR151 (trpC2, lys-3, metB10). These strains, as well as E. coli C600 $r_k^- m_k^-$, were routinely grown at 37°C in LB medium supplemented with the appropriate antibiotics. Selective concentrations were : ampicillin (Ap), 200 µg/ml; erythromycin (Em), 10 µg/ml; kanamycin (Km), 10 µg/ml. Plasmids constructed during this work are described in the text. They are all derived either from the E. coli replicon pUC4 (ampicillin-resistant; [8]), or the B. subtilis vectors pE194cop-6 (erythromycin-resistant; [9]), and pPL703 (kanamycin-resistant; [10]). Plasmid pCGV14, containing the 2.3-kb PstI fragment I of φ105 [11] inserted in the PstI site of pE194cop-6, has been described earlier [3].

φ105 infectivity assays

These were performed as described [7], using wild-type φ105 (10^9 pfu/ml). Strains scored as immune in this work show at least a 10^6-fold reduction in plaque-forming ability compared to sensitive strains.

General recombinant DNA methods

All procedures for plasmid and phage DNA preparation, restriction ligation, and transformation of competent B. subtilis or E. coli were as described previously [7].

Nuclease Bal31-induced deletion mutagenesis

Plasmid pCGV14 DNA (Fig. 3A) was linearized with HindIII, phenol-extracted and ethanol-precipitated, and subjected to Bal31 digestion essentially as described [12,13]. About 8 µg of DNA in 0.1 ml buffer was di-

gested with 1 unit of Bal31 at 30°C. Fractions were removed after 2.5, 5, 7.5, 12.5, and 15 minutes, and the extent of digestion was monitored by agarose gel electrophoresis. Subsequently, the appropriate fractions were pooled, phenol-extracted and ethanol-precipitated, followed by treatment with the Klenow fragment of DNA polymerase I. The resulting blunt-end fragments were self-ligated using T4 DNA ligase in the presence of 0.5 mM spermidine, prior to transformation of B. subtilis BR151. Individual erythromycin-resistant transformants were tested for phage φ105 infectivity, and their plasmid content was analyzed using BclI and PstI digestions to approximately determine the extent of the deletion. Subsequently, the deletion endpoints in a few of the pCGV14Δ plasmids (see Results) were determined at the nucleotide level.

DNA sequencing

Sequences were determined using the Maxam-Gilbert [14] method. Fragments were 5' end-labelled using γ-^{32}P-ATP and T4 polynucleotide kinase. For 3' end-labelling, the Klenow enzyme of DNA polymerase I was used.

RESULTS

The φ105 repressor is encoded by the promoter-proximal of two overlapping genes

We have previously shown [7] that the left part of φ105 EcoRI fragment F, more specifically the 1100-bp EcoRI-HindIII subfragment (Figure 1), most probably encodes a repressor function by virtue of its ability to confer immunity to φ105 infection when inserted into a B. subtilis cloning vector. We wanted to determine the nucleotide sequence of this region of the phage genome. To this end, the 3.2-kb φ105 EcoRI-F fragment and the adjacent 0.9-kb EcoRI-H fragment were each cloned in the E. coli replicon pUC4, yielding respectively plasmids pUC4/F and pUC4/H. A detailed restriction map of the repressor region, together with the adopted Maxam-Gilbert sequencing strategy is shown in Figure 1. Both strands of a 1306-bp segment, extending from an RsaI site in EcoRI-H to the HindIII site, were completely sequenced. An additional sequence run, using plasmid pCGV14 DNA, from the leftmost RsaI site in EcoRI-F to the PstI site in EcoRI-H confirmed the contiguity of both EcoRI fragments. Within our sequence the position of the PstI, SstI and NcoI sites is in accordance with the recently reported φ105 restriction map [11]. The sequence was analyzed for open-reading frames exceeding 300 nucleotides in length. Two partially overlapping open-reading frames were found, designated ORF1 and ORF2 in Figure 1. Both are

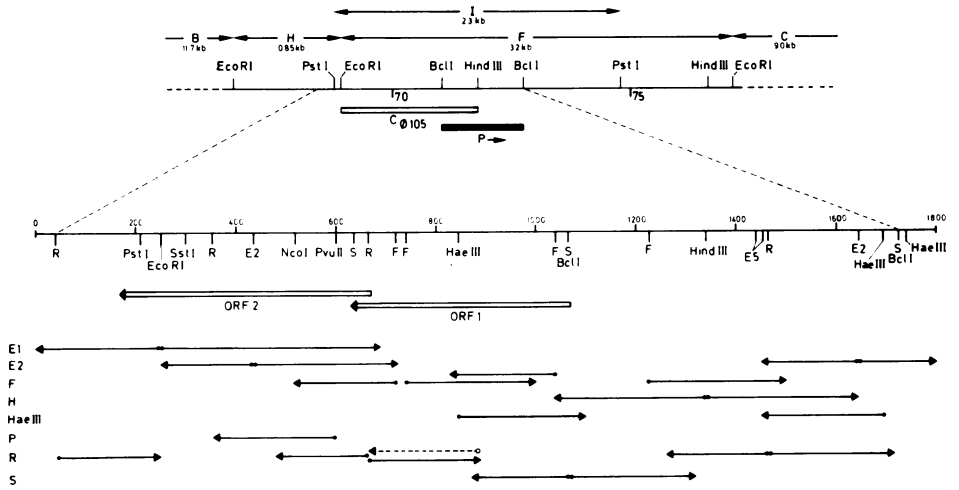


Figure 1. Map position and sequencing strategy of the ϕ105 repressor region.
The top shows a restriction map of a ϕ105 DNA segment around 70–75% genome length, including EcoRI fragments H and F. The location of PstI fragment I is also indicated. The open bar shows the map position of the repressor (Cϕ105). The black bar shows the position of the early promoter BclI fragment (P). The bottom shows the detailed restriction map of the repressor region. The position and direction of the two ORFs is indicated below the restriction map. Solid arrows indicate sequences derived by 5' end labelling. A broken line indicates a sequence generated by 3' end labelling. Distances are in bp. E1, EcoRI; E2, EcoRII (BstNI); E5, EcoRV; F, HinfI; H, HindIII; P, PvuII; R, RsaI; S, Sau3A.

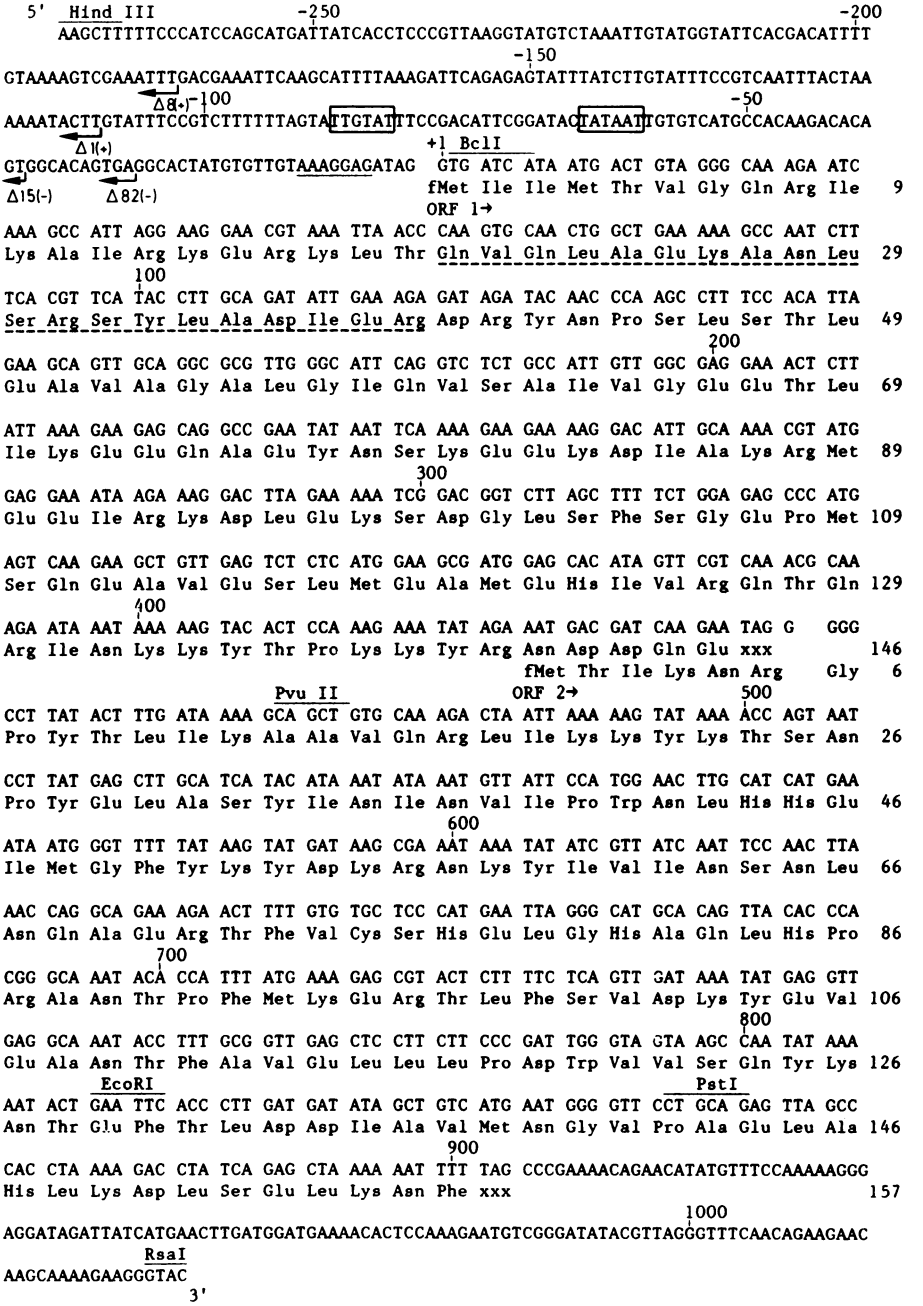
orientated from right to left on the conventional ϕ105 map. The nucleotide and deduced amino acid sequences of the ORF1-ORF2 gene pair are shown in Figure 2. Since ORF2 encompasses the EcoRI site and is thus not entirely contained within the EcoRI-HindIII fragment, ORF1 at this stage seemed to be the most likely candidate to encode the repressor function. However, the possibility remained that the C-terminal part of the polypeptide is dispensable for repressor activity, and hence, ORF2 could not yet be definitely excluded.

Figures 1 and 2 show the convenient location of a PvuII site with respect to the functional separation of both coding sequences : this site is well downstream from the stop codon of ORF1, whereas it occurs within the 14th codon from the N-terminus of ORF2. Hence, we decided to assay the 740-bp PvuII-HindIII fragment for repressor activity. The B. subtilis vector pPL703 [10] containing unique HindIII and PvuII sites, was digested

with both enzymes and ligated to HindIII+PvuII-digested pUC4/F. A large number of kanamycin-resistant B. subtilis transformants were subsequently tested for sensitivity to ϕ105. Only those cells containing a recombinant plasmid (designated pCGV23; see Fig. 3B) in which the 850-bp HindIII-PvuII vector fragment was substituted by the HindIII-PvuII fragment comprising ORF1, were found to be immune to ϕ105 superinfection. This finding conclusively demonstrated that ORF1 codes for the repressor.

Immediately preceding ORF1 is a sequence 5' AAAGGAG 3', showing a 7-bp complementarity to the 3' end of B. subtilis 16S rRNA [15], and hence representing a strong ribosomal binding site. In view of the frequent use of initiation codons other than ATG in B. subtilis [16], there are two possible candidates to serve as translational starts for the repressor : the GTG codon at position +1, and the in-frame ATG codon at position +10. We have as yet not been able to discriminate between these two possibilities. The respective distances between these codons and the above-defined Shine-Dalgarno sequence are 4 and 13 bases, whereas on the average a spacing of 8 to 9 nucleotides is found for B. subtilis genes [16,17]. Thus, the ϕ105 repressor monomer either consists of 146 or 143 amino acids, with calculated molecular weights of 16,745 or 16,387 daltons. These values correspond to a reasonable extent with the results obtained by examining the polypeptide synthesis pattern in B. subtilis minicells after infection with ϕ105 [18]. One of the earliest polypeptides, whose synthesis was not inhibited by chloramphenicol treatment, showed an apparent molecular weight of 18,000 on SDS-acrylamide gel electroforesis, and might therefore represent the product of this repressor gene.

Translation of ORF2 most probably starts at the ATG codon overlapping the triplets for residue 142 and 143 of the repressor. The function of the ORF2 gene product (157 amino acids) is unknown but its expression is obviously coregulated tightly with that of the repressor. This kind of translational overlap is quite unusual, but by no means unique [19]. In the case of the overlapping coat and lysis protein genes of RNA coliphages, it has been proposed that translation of the distal gene depends on a frameshift during translation of the proximal gene [20]. We have as yet no evidence that this might be true also for the repressor-ORF2 genes, but it is clear that the putative, if any, Shine-Dalgarno site preceding ORF2, 5' AAAGAAA 3', shows less complementarity (5 bp) to the 16S rRNA 3' end than its repressor counterpart.


5' Hind III
-250
-200
AAGCTTTTTCCCATCCAGCATGATTATCACCTCCCGTTAAGGTATGTCTAAATTGTATGGTATTCACGACATTTT
-150
GTAAAAGTCGAAATTTGACGAAATTCAAGCATTTTAAAGATTCAGAGAGTATTTATCTTGTATTTCCGTCAATTTACTAA
Δ8(+)
-100
-50
AAAATACTTGTATTTCCGTCTTTTTTAGTATTGTATTTCCGACATTCGGATACTATAATTGTGTCATGCCACAAGACACA
Δ1(+)
+1 BclI
GTGGCACAGTGAGGCACTATGTGTTGTAAAGGAGATAG GTG ATC ATA ATG ACT GTA GGG CAA AGA ATC
Δ15(-) Δ82(-)
fMet Ile Ile Met Thr Val Gly Gln Arg Ile 9
ORF 1→
AAA GCC ATT AGG AAG GAA CGT AAA TTA ACC CAA GTG CAA CTG GCT GAA AAA GCC AAT CTT
Lys Ala Ile Arg Lys Glu Arg Lys Leu Thr Gln Val Gln Leu Ala Glu Lys Ala Asn Leu 29
100
TCA CGT TCA TAC CTT GCA GAT ATT GAA AGA GAT AGA TAC AAC CCA AGC CTT TCC ACA TTA
Ser Arg Ser Tyr Leu Ala Asp Ile Glu Arg Asp Arg Tyr Asn Pro Ser Leu Ser Thr Leu 49
200
GAA GCA GTT GCA GGC GCG TTG GGC ATT CAG GTC TCT GCC ATT GTT GGC GAG GAA ACT CTT
Glu Ala Val Ala Gly Ala Leu Gly Ile Gln Val Ser Ala Ile Val Gly Glu Glu Thr Leu 69
ATT AAA GAA GAG CAG GCC GAA TAT AAT TCA AAA GAA GAA AAG GAC ATT GCA AAA CGT ATG
Ile Lys Glu Glu Gln Ala Glu Tyr Asn Ser Lys Glu Glu Lys Asp Ile Ala Lys Arg Met 89
300
GAG GAA ATA AGA AAG GAC TTA GAA AAA TCG GAC GGT CTT AGC TTT TCT GGA GAG CCC ATG
Glu Glu Ile Arg Lys Asp Leu Glu Lys Ser Asp Gly Leu Ser Phe Ser Gly Glu Pro Met 109
AGT CAA GAA GCT GTT GAG TCT CTC ATG GAA GCG ATG GAG CAC ATA GTT CGT CAA ACG CAA
Ser Gln Glu Ala Val Glu Ser Leu Met Glu Ala Met Glu His Ile Val Arg Gln Thr Gln 129
400
AGA ATA AAT AAA AAG TAC ACT CCA AAG AAA TAT AGA AAT GAC GAT CAA GAA TAG G GGG
Arg Ile Asn Lys Lys Tyr Thr Pro Lys Lys Tyr Arg Asn Asp Asp Gln Glu xxx 146
fMet Thr Ile Lys Asn Arg Gly 6
Pvu II
ORF 2→
500
CCT TAT ACT TTG ATA AAA GCA GCT GTG CAA AGA CTA ATT AAA AAG TAT AAA ACC AGT AAT
Pro Tyr Thr Leu Ile Lys Ala Ala Val Gln Arg Leu Ile Lys Lys Tyr Lys Thr Ser Asn 26
CCT TAT GAG CTT GCA TCA TAC ATA AAT ATA AAT GTT ATT CCA TGG AAC TTG CAT CAT GAA
Pro Tyr Glu Leu Ala Ser Tyr Ile Asn Ile Asn Val Ile Pro Trp Asn Leu His His Glu 46
600
ATA ATG GGT TTT TAT AAG TAT GAT AAG CGA AAT AAA TAT ATC GTT ATC AAT TCC AAC TTA
Ile Met Gly Phe Tyr Lys Tyr Asp Lys Arg Asn Lys Tyr Ile Val Ile Asn Ser Asn Leu 66
AAC CAG GCA GAA AGA ACT TTT GTG TGC TCC CAT GAA TTA GGG CAT GCA CAG TTA CAC CCA
Asn Gln Ala Glu Arg Thr Phe Val Cys Ser His Glu Leu Gly His Ala Gln Leu His Pro 86
700
CGG GCA AAT ACA CCA TTT ATG AAA GAG CGT ACT CTT TTC TCA GTT GAT AAA TAT GAG GTT
Arg Ala Asn Thr Pro Phe Met Lys Glu Arg Thr Leu Phe Ser Val Asp Lys Tyr Glu Val 106
800
GAG GCA AAT ACC TTT GCG GTT GAG CTC CTT CTT CCC GAT TGG GTA GTA AGC CAA TAT AAA
Glu Ala Asn Thr Phe Ala Val Glu Leu Leu Leu Pro Asp Trp Val Val Ser Gln Tyr Lys 126
EcoRI
PstI
AAT ACT GAA TTC ACC CTT GAT GAT ATA GCT GTC ATG AAT GGG GTT CCT GCA GAG TTA GCC
Asn Thr Glu Phe Thr Leu Asp Asp Ile Ala Val Met Asn Gly Val Pro Ala Glu Leu Ala 146
900
CAC CTA AAA GAC CTA TCA GAG CTA AAA AAT TTT TAG CCCGAAAACAGAACATATGTTTCCAAAAAGGG
His Leu Lys Asp Leu Ser Glu Leu Lys Asn Phe xxx 157
1000
AGGATAGATTATCATGAACTTGATGGATGAAAACACTCCAAAGAATGTCGGGATATACGTTAGGGTTTCAACAGAAGAAC
RsaI
AAGCAAAAGAAGGGTAC
3'

Deletion Analysis of the 5' Upstream Control Region of the Repressor-ORF2 Genes

Having defined the structural part of the φ105 repressor gene, we were next interested in identifying the control signals for the repressor-ORF2 transcription unit in the 5'-flanking DNA. Plasmid pCGV14 (Figure 3A) contains the intact ORF1 with over 1 kb of upstream sequence, and as expected it confers immunity to φ105 infection [3].

In this plasmid, the HindIII site 272 bp upstream of ORF1 is unique. To create a series of deletions starting from this site and progressing towards the beginning of ORF1, HindIII-linearized pCGV14 DNA was treated with the Bal31 nuclease as described in Materials and Methods.

As a result we obtained a collection of pCGV14 mutants (pCGV14Δ...) with deletions extending in both directions from the HindIII site, and ranging in size from less than 100 to over 2000 bp. In this collection the two phenotypic classes were represented : some pCGV14Δ plasmids had retained the superinfection immunity phenotype of the parent pCGV14 whereas in the majority it was lost. All plasmids with a deletion extending beyond the BclI site at the N-terminus of ORF1 fell into the latter category, indicating that an intact ORF1 is not only sufficient, but also necessary for superinfection immunity. Moreover, some deletion plasmids, such as pCGV14Δ15 and pCGV14Δ82, still contained this BclI site and yet gave rise to φ105-sensitive cells. Sequence analysis (Figure 2) showed that pCGV14Δ15 has retained 36 nucleotides upstream of the GTG, and pCGV14Δ82 ends at -26. The deletion endpoints in two immunity-conferring plasmids, pCGV14Δ1 and pCGV14Δ8, were localized at -109 and -182 respectively. These results clearly indicate that the region between -109 and -36 upstream from the GTG is required for the expression of superinfection immunity, even from a multicopy plasmid. We propose that the promoter for the repressor-ORF2 transcription unit is contained within this region. Figure 2 shows the presence of two hexanucleotides, 5' TTGTAT 3' and 5' TATAAT 3' separated by

Figure 2. Nucleotide and deduced amino acid sequences of the φ105 repressor (ORF1)-ORF2 genes and their flanking regions.
The sequence is shown in reversed orientation with respect to Figure 1, in order to display the sense strand. It extends from the HindIII site 272 bp upstream of the repressor-coding sequence, to an RsaI site within the φ105 EcoRI-H fragment. Nucleotide numbering starts from the GTG codon (+1). Amino acids are numbered at the right. The endpoint of the deletions in four PCGV14Δ... mutants is indicated. The promoter "-35" and "-10" regions are boxed, and the Shine-Dalgarno sequence is underlined. The dashed line in ORF1 indicates a putative DNA-binding region.

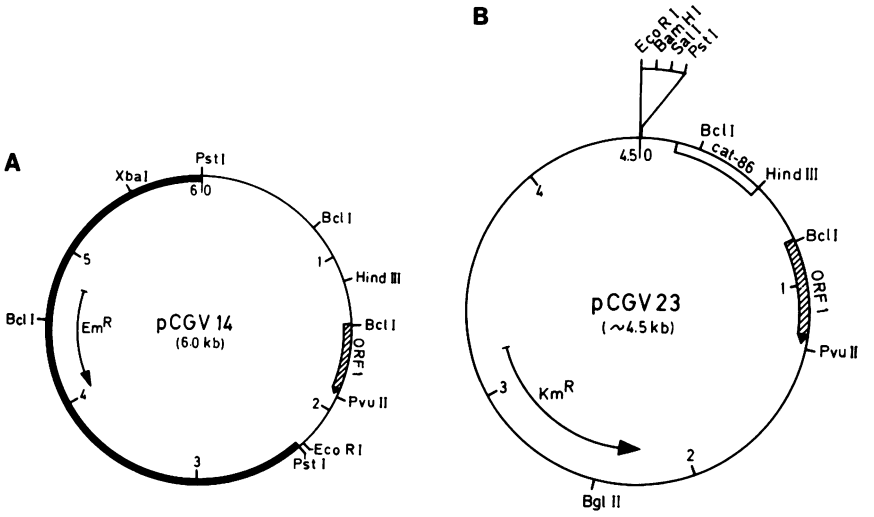


Figure 3.

A. Circular map of plasmid pCGV14 [3].
The orientation of the ϕ105 PstI-I fragment (thin line) with respect to pE194cop-6 (thick line) was determined by digestion with Em^R XbaI+PvuII. The position of ORF1 is indicated by the dashed arrow. Em^R, erythromycin resistance.

B. Circular map of pCGV23.
This plasmid is derived from pPL703 [10] through substitution of the 850-bp HindIII-PvuII vector fragment with the 740-bp HindIII-PvuII ϕ105 fragment comprising ORF1. Km^R, kanamycin resistance; cat-86, a chloramphenicol resistance gene originating from Bacillus pumilus.

exactly 17 bp. In view of the homology between these sequences and the consensus for a σ^{55}-dependent promoter [17], it is likely that these serve as "-35" and "-10" recognition sites for the major vegetative form of B. subtilis RNA polymerase. This is consistent with the idea that the repressor is probably one of the first phage genes to be expressed after infection, and therefore its transcription should rely entirely upon the host machinery.

We further note that these promoter elements are preceded by a very AT-rich region, (76% between positions -150 and -90), a feature common to several other B. subtilis genes [16,17].

The ϕ105 repressor polypeptide : properties and comparison with known repressors

The codon usage and amino acid compositions for ϕ105 repressor and ORF2, as deduced from the nucleotide sequence, are shown in Table I.

Table I. Codon usage for the φ105 repressor (ORF1) and ORF2 genes

	ORF1	ORF2		ORF1	ORF2		ORF1	ORF2		ORF1	ORF2
Phe TTT	1	5	Ser TCT	3	0	Tyr TAT	2	8	Cys TGT	0	0
TTC	0	2	TCC	1	2	TAC	3	1	TGC	0	1
Leu TTA	3	4	TCA	3	3	Och TAA	0	0	Umb TGA	0	0
TTG	1	2	TCG	1	0	Amb TAG	1	1	Trp TGG	0	2
Leu CTT	5	5	Pro CCT	0	3	His CAT	0	4	Arg CGT	4	1
CTC	1	1	CCC	1	1	CAC	1	2	CGC	0	0
CTA	0	4	CCA	2	3	Gln CAA	7	2	CGA	0	1
CTG	1	0	CCG	0	0	CAG	2	2	CGG	0	1
Ile ATT	6	2	Thr ACT	3	4	Asn AAT	4	11	Ser AGT	1	1
ATC	2	3	ACC	1	3	AAC	1	3	AGC	2	1
ATA	4	5	ACA	1	1	Lys AAA	9	9	Arg AGA	6	2
Met ATG	5	4	ACG	1	1	AAG	5	4	AAG	1	1
Val GTT	4	6	Ala GCT	2	2	Asp GAT	3	5	Gly GGT	1	1
GTC	1	2	GCC	4	1	GAC	4	1	GGC	3	0
GTA	1	2	GCA	4	7	Glu GAA	14	4	GGA	1	0
GTG	1	2	GCG	2	1	GAG	6	7	GGG	1	3
									Total :	146	157

In what follows, we have tentatively assigned translation initiation to the GTG codon and assume that the terminal residue of the mature repressor polypeptide is isoleucine (Figure 2). The repressor does not contain cysteine or tryptophan, but has a high content of glutamic acid (20 residues). Its pI is likely to be close to neutral, given the sum of acidic residues (Asp + Glu = 27), and basic residues (Arg + His + Lys = 26). The most striking overall property of the polypeptide is its pronounced hydrophilicity, as illustrated in the hydropathy profile (Figure 4). Very hydrophilic regions extend from residues 10–19, 36–41, 71–100 and 130–146, whereas the only hydrophobic regions are found between residues 1–5 and 51–64. Clearly, this feature refers to the cytoplasmic localization of the repressor protein.

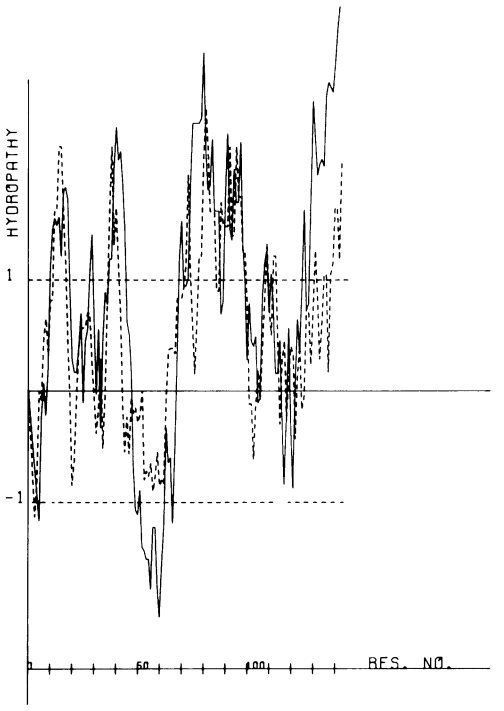


Figure 4. Hydropathy profiles of the ϕ105 repressor, according to the method of Hopp and Woods [31], solid line, and to that of Kyte and Doolittle [32], broken line.
Values were obtained using the average of a moving segment of 7 residues. Hydrophilic regions are found above the midline, hydrophobic are below.

In order to elucidate which region(s) of the polypeptide might be involved in DNA binding, we carried out an extensive comparison of its amino acid sequence with that of several other repressors functional in Gram-negative bacteria such as E. coli or S. typhimurium. These included : λ cI repressor [21,22], λ cro protein [23], phage P22 c2 repressor [24], lac repressor [25], gal repressor [26], and the lexA gene product [27]. The most favorable alignment for each pair of sequences was obtained using the Needleman-Wunsh algorithm [28], as modified by Kanehisa [29].

Table II summarizes the results for each pairwise comparison. The overall homology, at the primary structure level, between ϕ105 and any of the other repressors is insignificant as the highest proportion of identical residues, obtained with the lexA protein, amounts to less than 18%. However, figure 5 shows that there may exist some meaningful local homology

Table II. Comparison of the φ105 repressor amino acid sequence with that of other repressors

Protein	No. of compared positions	Matches	%
λ c_I	146	14	9.6
λ cro	65	4	6.2
P22 c2	146	24	16.4
lac R	147	14	9.5
gal R	146	16	11.0
lex R	152	27	17.8

in the N-terminal regions of φ105 repressor and phage P22 c2 repressor. On a total of 57 compared positions, 17 homologies are found without the need to introduce gaps. This could be relevant in view of our recent finding that phage φ105 contains two immunity regions [3]. We speculate that this sequenced gene corresponds to the immC region of phage P22, and that the other φ105 region (located in the EcoRI-B fragment) is analogous to the mnt/ant (or immI) region of P22.

A further comparison step was based on the well documented observation that all known repressors, regardless of the specific operator sequence to which they bind, contain a stretch of 20 contiguous residues exhibiting an α-helix – turn – α-helix configuration [30]. This three-dimensional structural element is essential for interaction with the major groove of the DNA helix. We examined the φ105 sequence for regions whose sequence would be compatible, at each crucial position, with the sequences of this DNA-binding domain in a number of other repressors [30]. As a result we propose the region extending from position 20 (Gln) to 39 (Arg) as the most likely

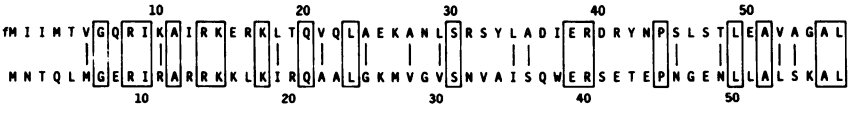


Figure 5. Alignment of the N-terminal amino acid sequences of φ105 repressor (top) and S. typhimurium phage P22 c2 repressor (bottom). The one-letter notation for amino acids is used. Residue numbers for φ105 repressor are identical to Figure 2. In P22 c2 repressor, the N-terminal methionine is deformylated, but not removed after translation [23]. Hence, numbering starts with this residue. Positions showing identical residues are boxed. Positions with a functionally rather conservative substitution are indicated by a vertical connector line.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Helix 2								turn			Helix 3								
λ cI	Q	E	S	V	A	D	K	M	G	M	G	Q	D	G	V	G	A	L	F	M
	33																			52
λ cro	Q	T	K	T	A	K	D	L	G	V	Y	Q	S	A	I	N	K	A	I	H
	16																			35
P22 c2	Q	A	A	L	G	K	M	V	G	V	S	N	V	A	I	S	Q	W	E	R
	21																			40
Lac R	L	Y	D	V	A	E	Y	A	G	V	S	Y	Q	T	V	S	R	V	V	N
	6																			25
Gal R	I	K	D	V	A	R	L	A	G	V	S	V	A	T	V	S	R	V	I	N
	4																			23
Lex R	R	A	E	I	A	S	R	L	G	F	R	S	P	N	A	A	E	E	H	L
	28																			47
γδ resolvase	A	S	H	I	S	K	T	M	N	I	A	R	S	T	V	Y	K	V	I	N
	161																			180
φ105 R	Q	V	Q	L	A	E	K	A	N	L	S	R	S	Y	L	A	D	I	E	R
	20																			39

Figure 6. Amino acid sequence homologies (one-letter notation) between several repressors in the region corresponding to the α2-turn-α3 configuration of λ cI and λ cro.
Except for the Φ105 repressor, data have been reproduced from [30]. The amino acid positions within each individual polypeptide are shown below its sequence. The crucial positions (residues 5, 9, 15, 18) are boxed.

candidate. In Figure 6 we have aligned this region with that of seven other repressors. The most important positions are 5, 9, 15 and 18 which should have alanine, glycine, and two hydrophobic amino acids, respectively. The φ105 sequence conforms well in three of these positions, whereas in position 9 it has asparagine which is, however, also found at that position in γδ-resolvase. Furthermore, hydrophobic residues should be present in positions 4, 8, and 19. Our φ105 sequence again agrees at two of these positions (4-Leu and 8-Ala), and the glutamic acid at 19 is also present in the P22 c2 repressor. Lastly, hydrophilic residues should occupy positions 1-3, 6-7, 11-14 and 16-17. Again, the φ105 sequence shows hydrophilic side chains in 9 out of these 11 positions. It is clear, however, that these arguments should be substantiated by further work, such as the isolation and mapping of φ105 repressor mutants affected in operator binding.

DISCUSSION

We have presented a detailed characterization of an immunity region of the B. subtilis temperate phage φ105, including the structural gene for the

repressor and the region controlling its transcription and translation. Throughout this work we used immunity to ϕ105 infection as a simple assay for the repressor function.

We should note, however, that a ϕ105 promoter-operator site controlled by the repressor has been localized within the 662-bp BclI fragment (Figure 1) immediately upstream from the repressor coding sequence [7]. This promoter directs transcription from left to right, and therefore can be topologically considered the equivalent of λ p_R with respect to the ϕ105 repressor. In addition, this promoter has been inserted in front of the cat-86 gene in plasmid pPGV10ϕ [7]. Using such a hybrid construction, we were able to assay the repressor function more directly, by testing the chloramphenicol resistance of B. subtilis [pPGV10ϕ] strains transformed with the different repressor plasmids constructed during this work, such as pCGV14 and its deletion derivatives. A perfect correlation was found between the superinfection immunity phenotype and the inhibition of cat-86 expression, resulting in Cm sensitivity of the double transformant. A ϕ105-sensitive phenotype, on the other hand, was always accompanied by the restoration of cat-86 expression, giving rise to $Em^R Cm^R$ cells (data not shown). Hence, we believe that we have characterized a true phage repressor in the full sense of the word, which is functionally analogous to, for example, the λ cI gene product.

There are, however, obvious differences in the organization of the ϕ105 immunity region compared to that of λ or the immC region of phage P22. First, the ϕ105 repressor, unlike its Gram-negative counterparts, forms part of a translationally overlapping pair of cistrons. The function of ORF2 as well as the reason why its expression should be coordinated so tightly with that of the repressor, are intriguing questions for which we have no answer to date.

Second, preliminary mapping data on the above mentioned target promoter-operator site [Dhaese et al., in preparation] indicate that it is located to the right of the HindIII site (Figure 1), and hence, at least 272 bp removed from the repressor initiation codon. In λ and P22, the p_R-O_R signals are immediately adjacent to the repressor coding sequence [1].

Our results suggest further that the ϕ105 repressor transcript, starting most likely some 10 bp downstream of the "TATAAT" hexanucleotide in Figure 2, contains at least 50 bases of untranslated leader, including a strong ribosomal binding site. Again, this is different from λ cI whose

transcription from the p_{RM} promoter, initiates at the ATG codon [33] and the P22 c2 maintenance RNA which contains only four 5'-untranslated nucleotides [34].

Using a comparative approach, we have tentatively identified a DNA-binding region within the φ105 repressor polypeptide, extending from residue positions 20 to 39. It is now possible to test this assignment by targeted mutagenesis of the corresponding DNA sequence.

In conclusion, we believe that this work only constitutes a modest first approach to unravel the possible intricacies of the genetic organization of the φ105 immunity system. Two important, unanswered questions are : (i) what is the number, position and structure of the operator sites; and (ii) does the φ105 repressor also regulate its own transcription ?

ACKNOWLEDGEMENTS

We are grateful to Ivo Zaenen and Pierre Lens for computer processing of the primary sequence data and for running the extensive compare programs. We further thank Marie-Reine Dobbelaere and Jan Gielen for expert technical assistance, Albert Verstraete and Karel Spruyt for photographical work and Martine De Cock for typing this manuscript. This research was supported by a joint IWONL-Industries contract (#4461).

[1]Present address: Plant Genetic Systems, Plateaustraat 22, B-9000 Gent, Belgium

*To whom correspondence should be addressed

REFERENCES

1. Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (1983) Lambda II (Cold Spring Harbor Monographs Series, Vol. 13), pp. 1-694, Cold Spring Harbor Laboratory, New York.
2. Rutberg, L. (1982). in The molecular biology of the bacilli, Dubnau, D. Ed., pp. 247-288, Academic Press, New York.
3. Dhaese, P., Dobbelaere, M.-R. and Van Montagu, M. (1985) Mol. Gen. Genet., in press.
4. Flock, J.-I. (1977) Mol. Gen. Genet. 155, 241-247.
5. Cully, D.F. and Garro, A.J. (1980) J. Virol. 34, 789-791.
6. Uhlén, M., Flock, J.-I. and Philipson, L. (1981). Plasmid 5, 161-169.
7. Dhaese, P., Hussey, C. and Van Montagu, M. (1984) Gene 32, 181-194.
8. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
9. Weisblum, B., Graham, M.Y., Gryczan, T. and Dubnau, D. (1979) J. Bacteriol. 137, 635-643.
10. Mongkolsuk, S., Chiang, Y.-W., Reynolds, R.B. and Lovett, P.S. (1983). J. Bacteriol. 155, 1399-1406.
11. Bugaichuk, U.D., Deadman, M., Errington, J. and Savva, D. (1984) J. Gen. Microbiol. 130, 2165-2167.
12. Legerski, R.J., Hodnett, J.L. and Gray, Jr. H.B. (1978) Nucl. Acids Res. 5, 1445-1464.

13. Maniatis, T, Fritsch, E.E. and Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual, pp. 1-545, Cold Spring Harbor Laboratory, New York.
14. Maxam, A. and Gilbert, W. (1980) in Nucleic Acids, Part I, (Methods in Enzymology, Vol. 65), Grossman, L. and Moldave, K., Eds., pp. 499-559, Academic Press, New York.
15. Murray, C.L., and Rabinowitz, J.C. (1982). J. Biol. Chem. 257, 1053-1062.
16. Doi, R.H. (1984). in Biotechnology and Genetic Engineering Reviews, Vol. 2, Russell, G.C., Ed., pp. 121-155, Intercept, Newcastle upon Tyne.
17. Moran, C.P., Lang, N., Le Grice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, L. and Losick, R. (1982) Mol. Gen. Genet. 186, 229-346.
18. Reeve, J.N. (1977). Mol. Gen. Genet. 158, 73-79.
19. Normark, S., Bergström, S., Edlund, T., Grundström, T., Jaurin, B., Lindberg, F.P., and Olsson, O. (1983). Ann. Rev. Genet. 17, 499-525.
20. Kastelein, R.A., Remaut, E., Fiers, W., and Van Duin, J. (1982). Nature (London) 295, 35-41.
21. Sauer, R.T. (1978) Nature (London) 276, 301-302.
22. Sauer, R.T. and Anderegg, R. (1978) Biochemistry 17, 1092-1100.
23. Hsiang, M.W., Cole, R.D., Takeda, Y. and Echols, H. (1977) Nature (London) 270, 275-277.
24. Sauer, R.T., Pan, J., Hopper, P., Hehir, K., Brown, J. and Poteete, A.K. (1981) Biochemistry 20, 3591-3598.
25. Beyreuther, K. (1978) Nature (London) 274, 767.
26. von Wilcken-Bergmann, B. and Müller-Hill, B. (1982) Proc. Natl. Acad. Sci. USA 79, 2427-2431.
27. Horii, T. Ogawa, T. and Ogawa, H. (1981) Cell 23, 689-697.
28. Needleman, S.B. and Wunsch, C.D. (1970). J. Mol. Biol. 48, 443-453.
29. Kanehisa, M.I. (1982). Nucl. Acids Res. 10, 183-196.
30. Pabo, C.O. and Sauer, R.T. (1984). Ann. Rev. Biochem. 53, 293-321.
31. Hopp, T.P. and Woods, K.R.L. (1981) Proc. Natl. Acad. Sci. USA 78, 3824-3828.
32. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
33. Ptashne, M., Backman, K., Humayun, M.Z., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, R.T. (1976) Science 194, 156-161.
34. Poteete, A.R. and Ptashne, M. (1982) J. Mol. Biol. 157, 21-48.