Volume 13 Number 2 1985 Nucleic Acids Research

The nucleotide sequences of the tail fiber gene 36 of bacteriophage T2 and of genes 36 of the T-even type *Escherichia coli* phages K3 and Ox2

Isolde Riede, Klaus Drexler and Marie-Luise Eschbach

Max-Planck-Institut für Biologie, Correnssstrasse 38, D-7400 Tübingen, FRG

Received 24 October 1984; Revised and Accepted 31 December 1984

ABSTRACT

Genes 36 have been cloned from phage T2 and the T-even type phages K3 and Ox2. The products of these genes are part of the long tail fibers of the phages, they form the proximal moiety of the distal half fiber. The genes have been sequenced, the nucleotide sequence of gene 36 of phage T4 is known (Oliver, D.B. & Crowther, R.A. (1981) J.Mol.Biol. 153, 545-568). Comparison of the deduced amino acid sequences of the four proteins revealed a surprising pattern. These sequences can be divided into two highly conserved and one very variable region. The former consist of about 60 NH_2-terminal and 70 CO_2H-terminal residues flanking the variable middle region of about 100 residues. Thus, an identical and unique morphology can be formed by a number of different primary structures. It is proposed that the conserved areas are involved in binding of the proteins to the neighboring products of genes 35 and 37 and that this function has put constraints on the variability of the primary protein structure. The overall amino acid composition of the proteins is rather similar; the codon usage is that known for phage T4. The intercistronic region between genes 35 and 36 consisting of 62 base pairs and containing a presumed terminator for g35 transcription and the 'late type' promoter for transcription of genes 36, 37, and 38, is almost completely identical in the four phages.

INTRODUCTION

Most likely all T-even type phages recognize the bacterial surface with their long tail fibers. The composition of these fibers of phage T4 is well known. The proximal half of the fiber consists of a dimer of gene product (gp) 34 which is linked by one copy of gp35 to the distal half of the fiber which is composed of dimers of gp36 and gp37. The latter dimer is present at the distal part of this half fiber; the CO_2H-termini of the two polypeptides are located at or near the free end of the fiber (1). Genes 34-37 are closely linked and arranged in that order (2).

Besides the classical phages T2, T4, and T6 a number of morphologically identical phages were isolated, among them a group which uses the E.coli outer membrane protein OmpA as a receptor (3). Because of the interesting immunological cross-reactions between these phages, indicating a high degree of polymorphism of gp37 (3), the protein containing the receptor recognition area, we had cloned this gene of phage T2 and the OmpA-specific phage K3 (4). In both cases gene (g) 36 was also recovered. Using parts of the cloned g36 and g37 of phage K3 to probe restriction digests of a number of other phages it became evident that the polymorphism of their genes 37 is larger than expected; their genes 36, however, appeared to be rather conserved within a group of phages. For example, at least part of g36 of another OmpA-specific phage, 0x2, was fairly homologous to that of phage K3, while the genes 37 of the two phages screened as being very different. For this reason, and as described here, g37 of phage 0x2 has also been cloned and again its g36 was also obtained. Surprisingly therefore, a coevolution of the two genes does not seem to have occurred although their products are neighbors in a rather unique structure. Hybridization data do not, of course, really prove such a conclusion, it would require knowledge of the DNA sequence. With genes 36 of three different phages and one known sequence (g36 of phage T4 (5)) at hand we decided to determine the structure of the three genes not only for the evolutionary implications; comparison of such sequences should eventually allow to define rules for the architecture and assembly of such fibers.

MATERIALS AND METHODS

Bacteria and plasmids

Plasmids were maintained in the E.coli K12 strains JM103 (6) or F^-Z^- M15 (7) and purified according to Maniatis et al. (8). As vectors served plasmid pUC8 (9) or phages M13 mp8 and mp9 (10). The latter was propagated on strain JM103, all other phages on strain P400 (11). Cells were grown at 37° in L-broth supplemented with ampicillin (20 µg/ml) when required.

Nucleic acid techniques

DNA fragments were transferred from agarose gels to nitrocellulose according to Southern (12) and hybridized with probes labeled with ^{32}P as described (13). DNA sequencing was performed employing the dideoxy chain-terminating procedure (14) and the universal primer or reverse primer (New England Biolabs).

Cloning and sequencing of tail fiber genes of phage 0x2

The 0x2 genes were cloned into plasmid pUC8 (9) exactly as described for those of phage T2 (4). For detecting clones containing g36 and g37 a probe from phage K3 was used (stemming from plasmid pTU K3-32 (4)) which contained the end of g36 and the beginning of g37 and which was known to hybridize well with fragments of restricted 0x2 DNA. Several independent isolates proved to harbor plasmid pTU 0x2-15 (Fig.1). The TaqI-HindIII or TaqI-PstI fragments of this plasmid were subcloned into plasmid pUC8 to yield pTU 0x2-15H and pTU 0x2-15P, respectively. Insert DNA of these plasmids was sequenced as indicated in Fig.2 using the appropriate primers. The 0.4 Kb PstI fragment from pTU 0x2-15H, containing the end of g36 and the beginning of g37 (Figs.1 and 2) was cloned into phage M13 mp8. Both orientations were found and sequenced. The PstI site was overlapped by subcloning the righthand EcoRI-TaqI fragment of pTU Ox2-15 into M13 mp9.

Subcloning and sequencing of tail fiber genes of phages T2 and K3 (Fig.2)

Plasmid pTU T2-51H, harboring the end of g35, g36, and most of g37 of phage T2 has been described previously (4). From it a HindIII-XbaI and a HindIII-SmaI fragment were cloned into pUC8 and were sequenced using the universal primer or the reverse primer. The missing areas to the right of XbaI and SmaI were obtained by cloning XbaI-EcoRI and SmaI-EcoRI (the EcoRI site being in g37) fragments into pUC8.

Plasmid pTU K3-32B contains part of g34, g35, and part of g36 of phage K3; pTU K3-32H harbors the end of g35, g36 and part of g37 of this phage (4). With pTU K3-32H and pTU K3-32B the complete sequences of the HindIII-BamHI fragment containing the end of g35 and a part of g36 were

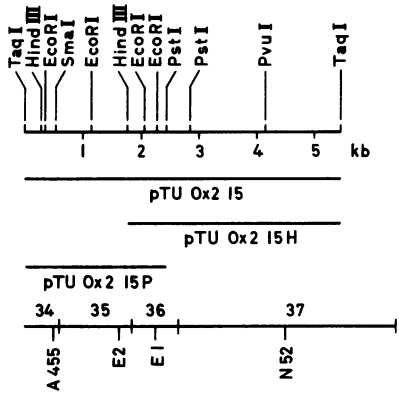


Fig.1: Organization of tail fiber genes of phage 0x2 on cloned fragments and the corresponding restricton map. pTU 0x2-15H and pTU 0x2-15P are HindIII and PstI subclones, respectively, from pTU 0x2-15. The T4 genes (bar at the bottom) are drawn on the same scale. The positions of the T4 amber mutants used for marker rescue tests are indicated as described (5,15,16).

obtained. With a new subclone of pTU K3-81, containing a BamHI fragment with the end of g36 and part of g37, the sequence of the missing part was determined. The overlap of the BamHI site was determined by a HaeIII subclone in M13 mp9.

Computer analysis of nucleotide and amino acid sequences

The program ALIGN (17) was used to evaluate the statistical significance of the homologies between the genes, i.e., to construct Fig.3. By changing several parameters (e.g., penalty for breaks) conditions were chosen yielding the highest alignment score with a minimum of breaks in the sequences. The program RELATE (17) was run to establish the genetic distances of the genes (intergene comparison) and to determine whether repetitive sequences exist within one gene (intragene comparison). A fragment length of 25 amino acid residue proved to be optimal for obtaining high values of SD units (Table 3); other fragment lengths gave similar results. The probability to obtain by chance SD units of 10,3, and 1 is 0.762×10^{-23}, 0.133×10^{-2}, and 0.159, respectively (17).

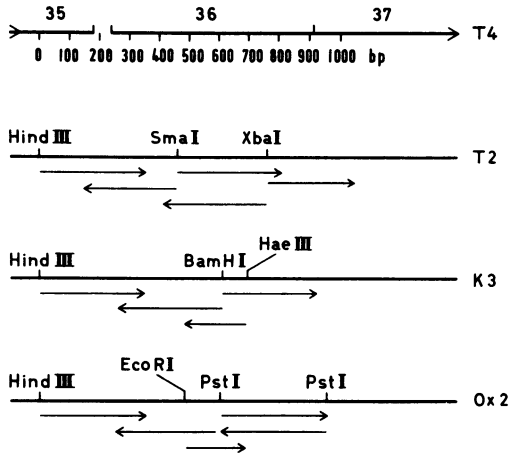
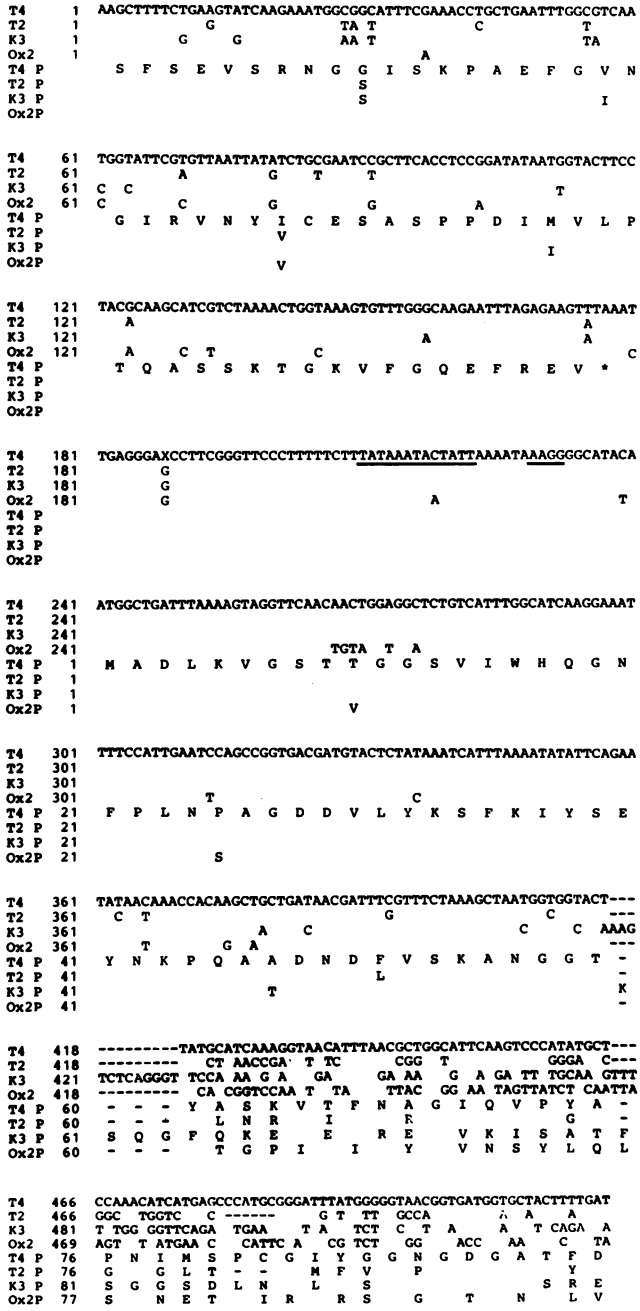


Fig.2: DNA sequencing strategies. The area of interruption within the upper arrow, indicating the direction of transcription of the T4 genes, represents the promoter area for genes 36, 37, and 38 (2). DNA sequences derived from only one strand did not show any ambiguities.

RESULTS

Clones containing g36 of phages 0x2, K3, and T2

Marker rescue tests showed that plasmid pTU 0x2-15 (Fig.1) could rescue the T4 amber mutants A455 (g34), E2 (g35), E1 (g36), but not N52 (g37). Considering the size of the cloned fragment and the known localization of N52 (5) there is little question that the corresponding area is present on the 0x2 DNA. The inability of pTU Ox-15 to rescue mutant N52 is certainly due to the heterology of the two genes 37: a probe from phage T4 containing g36 and most of g37 (3) hybridized almost only with the area of the cloned 0x2 DNA which is to the left of the PstI site at 2.8 Kb (Fig.2; data not shown). The localization of the other T4 markers in the subclones in plasmids pTU Ox2-15H and pTU Ox2-15P is shown in Fig.1. Plasmids containing genes 36 from phages K3 and T2 have been described recently (4). The strategies of sequencing of all three genes 36 are shown in Fig.2.



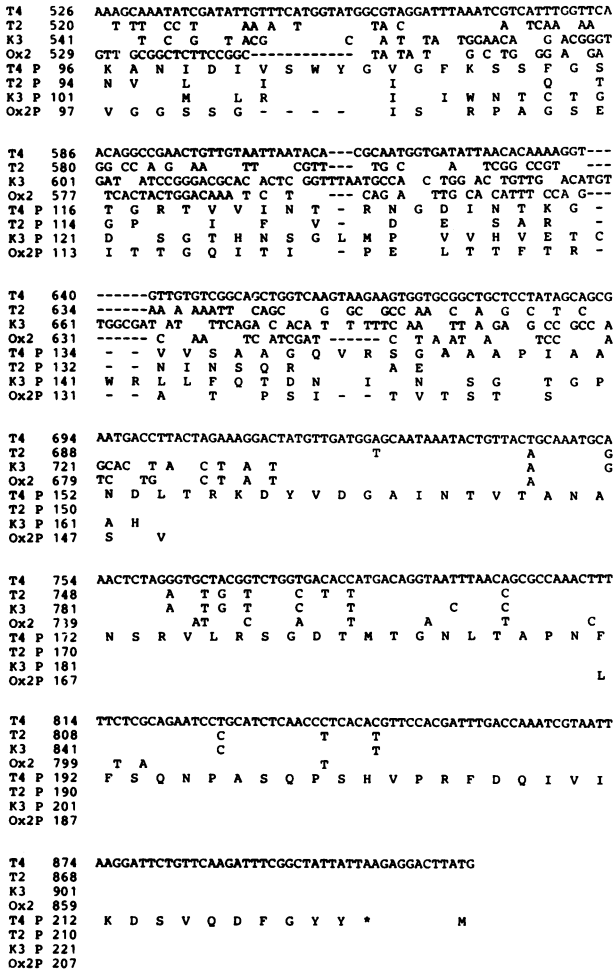


Fig.3: Nucleotide and deduced amino acid sequences. The nucleotide sequence of the end of g35 and of g36 (start of translation at bp241) of phage T4 (5) is compared with those of phages T2, K3, and Ox2. Only those bases and amino acid residues which differ from the T4-sequence are shown. The arrangement of the amino acid sequences in the area of heterology was arrived at by employing the computer program ALIGN (17). In most cases the best fit between all four sequences could be obtained. In a few instances best alignment of all sequences was not possible. In such doubtful cases the best fit with the sequence of phage T4 was chosen. At position 188 of the published sequence of T4 the symbol 8 (here given as X) stands for an ambiguous gel reading of G or C (R.A. Crowther, personal communication). The "juke box" for late transcription initiation (18) and the Shine-Dalgarno sequence are underlined.

Fig.4: Potential secondary structure of the transcriptional terminator of genes 35. The stop codon of gp35 is underlined.

Comparison of the DNA sequences of genes 36, and the corresponding deduced protein sequences from phages T4, T2, K3 and 0x2

The sequences are displayed in Fig.3. In those from phages K3, 0x2 and T2 two open reading frames were found. Comparison with sequences from tail fiber genes from phage T4 (5) showed that in all cases the first frame obviously represents the end of g35; it is highly conserved in all four phages. Similarly, the intercistronic region between genes 35 and 36, containing the terminator for g35 (Fig.4) and the promoter for the polycistronic mRNA for g36, g37, and g38 of phage T4 (2) is almost identical in all phages. Near identity continues from the beginning of g36 over a stretch of the gene encoding about 60 NH_2-terminal amino acid residues. A very heterologous area coding for about 100 residues follows which is fairly unique for each phage. The last part of the four genes, coding for about 70

Table 1: Codons used in the mRNA of tail fiber genes. The values for g36 of T4, T2, K3, Ox2 and g37 of T4 are given in that order. Codons corresponding to the known anticodons of tRNA's (19) encoded by phage T4 are underlined.

UUU	F	8:	8:	7:	6:32	UCU	S	6:	8:12:	10:28	UAU	Y	10:	8:	7: 6:29	UGU	C	0:	0:	2:	0: 1
UUC	F	3:	2:	4:	2:12	UCC	S	0:	0:	2:	2: 5	UAC	Y	0:	1:	1:	3: 9	UGC	C	1:	0:	0:	0: 1
UUA	L	2:	2:	7:	4:21	UCA	S	8:	6:	7:	6:20	UAA	*	0:	0:	0:	0: 0	UGA	*	0:	0:	0:	0: 0
UUG	L	1:	3:	3:	2:13	UCG	S	3:	2:	1:	1: 5	UAG	*	0:	0:	0:	0: 0	UGG	W	2:	2:	4:	1:10
CUU	L	1:	2:	0:	3: 7	CCU	P	2:	1:	0:	4:13	CAU	H	1:	2:	3:	1:13	CGU	R	0:	4:	3:	2:18
CUC	L	1:	1:	1:	1: 2	CCC	P	1:	3:	2:	1: 1	CAC	H	1:	0:	2:	1:11	CGC	R	1:	0:	0:	1:11
CUA	L	1:	1:	0:	0: 3	CCA	P	8:	6:	7:	7: 8	CAA	Q	7:	7:	6:	6:29	CGA	R	2:	3:	2:	2: 1
CUG	L	0:	1:	1:	0: 2	CCG	P	0:	1:	1:	0: 6	CAG	Q	1:	3:	3:	2: 8	CGG	R	1:	1:	0:	0: 0
AUU	I	7:	6:	6:	9:44	ACU	T	7:	6:	5:	13:49	AAU	N	11:	13:	8:	13:43	AGU	S	1:	0:	0:	4:21
AUC	I	3:	3:	1:	2:13	ACC	T	1:	2:	3:	4: 9	AAC	N	8:	7:	10:	3:35	AGC	S	1:	0:	0:	0:13
AUA	I	3:	5:	2:	3:13	ACA	T	7:	5:	6:	6:25	AAA	K	8:	6:	8:	6:31	AGA	R	2:	2:	3:	1:17
AUG	M	3:	3:	4:	2:16	ACG	T	0:	0:	3:	1: 6	AAG	K	3:	2:	3:	1: 8	AGG	R	1:	1:	0:	2: 3
GUU	V	8:	10:	8:	8:34	GCU	A	10:	4:	6:	5:42	GAU	D	11:	11:	10:	9:42	GGU	G	14:	11:	9:	12:65
GUC	V	2:	2:	1:	1: 5	GCC	A	1:	6:	1:	1:10	GAC	D	5:	5:	6:	3:13	GGC	G	5:	5:	10:	5:28
GUA	V	7:	3:	5:	8:17	GCA	A	8:	5:	5:	8:25	GAA	E	1:	3:	6:	4:25	GGA	G	4:	7:	7:	6:32
GUG	V	2:	2:	2:	0: 2	GCG	A	3:	4:	2:	1: 5	GAG	E	0:	0:	0:	0: 5	GGG	G	2:	2:	2:	1:11

Table 2: Predicted amino acid compositions of the gene products. The data are given in percent in order to take into account the differences in molecular weights.

Phage		T4	T2	K3	Ox2	T4
Gene No.		36	36	36	36	37
No. of residues		221	219	230	216	1026
predicted mol. wt.		23 350	23 500	24 750	22 700	109 200
Composition	A	10.0 %	8.7 %	6.1 %	6.9 %	8.0 %
	R	3.2	5.0	3.5	3.7	4.9
	N	8.6	9.1	7.8	7.4	7.6
	D	7.2	7.3	7.0	5.5	5.4
	C	0.5	0.0	0.9	0.0	0.2
	Q	3.6	4.6	3.9	3.7	3.6
	E	0.5	1.4	2.6	1.9	2.9
	G	11.3	11.4	12.2	11.1	13.3
	H	0.9	0.9	2.2	0.9	2.3
	I	5.9	6.4	3.9	6.5	6.8
	L	2.7	4.6	5.2	4.6	4.7
	K	5.0	3.7	4.8	3.2	3.8
	M	1.4	1.4	1.7	0.9	1.6
	F	5.0	4.6	4.8	3.7	4.3
	P	5.0	5.0	4.4	5.6	2.7
	S	8.6	7.3	9.6	10.7	9.0
	T	6.8	5.9	7.4	11.1	8.7
	W	0.9	0.9	1.7	0.5	1.0
	Y	4.5	4.1	3.5	4.2	3.7
	V	8.6	7.8	7.0	7.9	5.7

Table 3: Relationship of the genes 36. (A), the values indicate segment comparison scores (SD units) obtained by running the program RELATE (17). Upper numbers: results of intra-sequence comparisons of the proteins. Segments 25 amino acid residues long were compared (500 random runs). The scoring matrix was the mutation data matrix. Lower numbers: results of intra-sequence comparisons of the DNAs. Segments 75 nucleotides long were compared (100 random runs). The scoring matrix was the unitary matrix. (B), inter-sequence comparisons; the conditions of scoring and the arrangement of the values are the same as in A. (C) SD units of B in percent of the T4-T2 comparisons which were set 100%. (43 of 221 amino acid residues are different or missing (upper numbers) and 136 of 663 nucleotides are different or missing (lower numbers) in comparing T2 with T4).

B \ C / A	T4	T2	K3	Ox2
T4	-1.485 1.563	100 % 100 %	73.8 % 71.9 %	75.8 % 70.9 %
T2	50.692 85.476	-1.070 2.009	80.7 % 70.4 %	68.7 % 63.6 %
K3	37.411 61.422	40.905 60.207	0.883 0.738	66.6 % 65.3 %
Ox2	38.425 60.601	34.806 54.397	33.769 55.838	0.198 2.676

residues is again highly conserved and, as it is the case for the proximal parts of the genes, most of the base pair substitutions are silent. The codon usage (Table 1) is very similar in all four phages and also to that for g37 of T4 (5), suggesting that T2, K3 and Ox2 code for the same t-RNA species as does T4 (19).

It is interesting that despite the considerable heterologies of the genes the amino acid composition of the four genes 36 and that of gene 37 of phage T4 are rather similar (Table 2). Particularly striking is the high glycine content of all proteins. We also noted that the sum of the most hydrophobic residues (I, V, L, F, C, M, A) is remarkably constant (g36 of T4, 34.1%; of T2, 33.5%; of K3, 29.6%; of Ox2, 30.5%; g37 of T4, 31.3%).

Computer analysis of nucleotide and amino acid sequences

Segment comparison scores (17) of the four genes and proteins are shown in Table 3. Negative values of the segment comparison scores (intragene comparisons (A)) clearly show (T2 and T4) that the proteins do not contain repetitive sequences. Values lower than 1 are not significant, thus the proteins of K3 and Ox2 do also not consist of repetitive segments. Because of the use of different data matrices the values for SD units obtained from nucleotide sequence comparisons (intergene comparisons (B)) are higher than those obtained from amino acid sequence comparisons. Therefore, the segment comparison scores were normalized as percent of the scores obtained for the comparison of T2 with T4, the latter arbitrarily set 100%. The values for amino acid sequence comparisons are higher than those obtained for nucleotide sequences, suggesting that the former are more conserved than the latter.

Using the data of Table 3 we have attempted to construct a genealogical tree for the four phages but it was not possible to draw unequivocal conclusions. T2 and T4 clearly are the closest relatives; it is not obvious, however, where to place the other two phages. The degree of relatedness in terms of SD units, obtained from the segment comparison scores, between Ox2 and T4 are similar to those between K3 and T2 or K3 and T4, but the degree of homology between Ox2 and T2 is much lower.

DISCUSSION

Among the phages studied the gene products 36 can clearly be divided into three segments. Two, consisting of about 60 NH_2-terminal and about 70 CO_2H-terminal amino acid residues are highly conserved. The middle segments of the polypeptides, comprising about 100 residues, are rather unique for each phage. Almost certainly the two ends of the dimeric protein serve to bind to the neighboring gp35 and the dimer of gp37, and it would appear that these protein-protein interactions have put severe constraints on the variability of the corresponding primary structures. In line with this interpretation we have since found that the ends of g34, beginnings and ends of g35, and the beginning of g37 of the four phages are also highly conserved. Cysteine residues are absent from the conserved regions, thus binding of the dimeric gp36 to its neighbors does not involve disulfide bridges. Since tail fiber integrity is insensitive to large variations of salt concentration the binding of the proteins may be of hydrophobic nature.

The highly variable area of gp36 probably has no other functions than to establish the structure of a fiber. As pointed out before and concerning parts of gp37 of phages T4 and K3 (4), it certainly is remarkable that identical and unique morphologies can be formed by such different primary structures. For gp36 at least one other primary structure is possible. We have shown that DNA's of other T-even type phages do not hybridize with probes from T4 or K3 representing the area of g36 (3,4). Also, no serological cross-reactions were found in these cases (3).

The intercistronic region between g35 and g36 is almost completely conserved in all four phages. It contains a putative terminator for g35 transcription (Fig.4), its sequence is very similar to those for the late gene 23 and the quasilate gene 32; it differs by only one base from the terminator for g23 of phage T4 (18). This extreme conservation lends further credit to the proposed role of this sequence. The following sequence, TATAAATACTATT, is identical with the consensus sequence for late T4 promoters (18,20). The perfect homology

of the areas between terminator, promoter, Shine-Dalgarno sequence and translation initiation codon of this intercistronic region, particularly in view of the variability found in g36, points to important roles also of these sequences.

ACKNOWLEDGEMENTS

We are grateful to B. Foertsch for introducing us to the computer analyses and performing these at the facilities of the Max-Planck-Institut für Biochemie. We thank U. Rüther for plasmids and strains, E. Kellenberger and W.B. Wood for T4 amber mutants, and U. Henning for support and help in the preparation of the manuscript.

REFERENCES

1 Beckendorf, S.K. (1973) J.Mol.Biol. 73, 37-53.
2 King, J. and Laemmli, U.K. (1971) J.Mol.Biol. 62, 465-477.
3 Schwarz, H., Riede, I., Sonntag, I. and Henning, U. (1983) EMBO J. 2, 375-380.
4 Riede, I., Eschbach, M.-L. and Henning, U. (1984) Mol.Gen. Genet. 195, 144-152.
5 Oliver, D.B. and Crowther, R.A. (1981) J.Mol.Biol. 153,545-568.
6. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
7 Rüther, U., Koenen, M., Otto, K. and Müller-Hill, B. (1981) Nucleic Acids Res. 9, 4087-4098.
8 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual, Publ. Cold Spring Harbor Laboratory Press, New York.
9 Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
10 Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
11 Skurray, R.A., Hancock, R.E.W. and Reeves, P. (1974) J.Bacteriol. 119, 726-735.
12 Southern, E.M. (1975) J.Mol.Biol. 98, 503-518.
13 Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc.Natl.Acad. Sci. USA 76, 3683-3687.
14 Sanger, F. Nicklen, S. and Coulson, A.R. (1977) Proc.Natl. Acad.Sci. USA 74, 5463-5467.
15 Revel, H.R. (1981) Mol.Gen.Genet. 182, 445-455.
16 Beckendorf, S.K., Kim, J.S. and Lielausis, I. (1973) J.Mol. Biol. 73, 17-35.
17 Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Meth. Enzymol. 91, 524-545.
18 Christensen, A.C. and Young, E.T. (1983) in: Bacteriophage T4 (eds. C.K. Mathews, E.M. Kutter, G. Mosig, P.B. Berget, American Society for Microbiology, Washington, D.C.) pp. 184-188.
19 Schmidt, F.J. and Apirion, D. (1983) in: Bacteriophage T4 (eds. C.K. Mathews, E.M. Kutter, G. Mosig, P.B. Berget, American Society for Microbiology, Washington, D.C.) pp.208-217.
20 Elliott, T. and Geiduschek, E.P. (1984) Cell 36, 211-219.