Nucleotide sequence of the bacteriophage T4 gene 57 and a deduced amino acid sequence

Richard Herrmann

Mikrobiologie, Universität Heidelberg, Im Neuenheimer Feld 230, 6900 Heidelberg, FRG

SUMMARY
    A 693 basepair cloned fragment of bacteriophage T4 DNA, which
supports specifically growth of T4 amber mutants in gene 57, has
been sequenced. A polypeptide can be deduced from this sequence,
that is either 54 or 60 amino acids long depending which of two
AUG codons, 18 nucleotides apart, are used for initiation. The
size of this deduced polypeptide is compatible with the size of
a single polypeptide (based on polyacrylamide gel electrophore-
sis) synthesized *in vivo* in *E. coli* under the direction of the
cloned T4 DNA fragment.

INTRODUCTION

    The product of the bacteriophage T4 gene 57 is essential for
morphogenesis of both the tail fiber and the baseplate of phage
T4 (for a review on gene 57 in T4 morphogenesis see refs. 1 and
2). A single amber mutation in gene 57 causes loss of the two
phage-specific proteins gp57A ($\sim$ 6000 daltons) and gp57B ($\sim$18000
daltons). Amber mutations in gene 57 can be complemented by a
cloned, about 700 bp long, fragment of phage T4 DNA. This cloned
DNA fragment directs in *E. coli* the synthesis of a single poly-
peptide, that appears to be identical with gp57A, indicating
that this polypeptide is the only essential gene product missing
in gene 57 amber mutants and that gp57B is not essential for
morphogenesis of these mutants (3). Functionally, gp57A belongs
to a class of nonstructural, accessory proteins, that mediate
assembly of structural proteins into a more complex organisation
(1).

    In order to further characterize gene 57, the cloned 700 bp
T4 DNA fragment was sequenced. The nucleotide sequence should
not only provide the primary structure of gp57A, it should also
help to explain genetic data, e.g. the finding that all gene 57

amber mutants appear to represent mutations at a single site (2) and the observation that two phage proteins are lost as a result of single amber mutations in gene 57 (3).

MATERIALS AND METHODS
Source of DNA.

   DNA for sequencing derived either from plasmid pCG57-7 or from the single stranded DNA phages fd57-7 and fd57-71 (3).The plasmid and the two phages are clones which carry the same re-striction endonuclease TaqI generated DNA fragment (about 700 bp) of phage T4 DNA. pCG57-7 is a derivative of plasmid pBR322; fd57-7 and fd57-71 are derivatives of fd109-2 (3), which differ only with respect to the orientation of integration of the clo-ned T4 DNA fragment. Procedures for isolation of recombinant fd phages, single stranded phage DNA, double stranded fd DNA and plasmid DNA were as described (4, 5).
Preparation of DNA for sequencing.

   Heteroduplex formation between single stranded DNA's (4). The two single stranded DNA's (0.5 - 2 pmol single stranded DNA from 1 ml of culture) of fd57-7 and fd57-71 are a convenient source of double stranded T4 DNA insert, since a mixture of both DNA's form a heteroduplex in the T4 moiety of the DNA. To this end single stranded DNA's of both phages were dissolved in equimolar amounts in hybridization buffer (50 mM Tris HCl pH8, 1 mM EDTA, 100 mM NaCl) at a concentration of 1000 µg/ml and sealed in a glass capillary. The mixture was incubated for 1 min at 100° C, transferred to an 80° C water bath and allowed to cool to room temperature over a period of about 3 hours. 300 µg hybridized DNA was desalted by two ethanol precipitations, taken up in S1 buffer (6) and treated with 115 units S1 nuclease for 30 min at 20° C and 15 min at 45° C. DNA was again phenol extracted and purified from residual small DNA fragments by chromatography on Sephadex G150 in a pasteur pipette column using 10 mM $(NH_4)HCO_3$ as elution buffer. Heteroduplex formation and S1 digestion were al-ways checked by agarose gelelectrophoresis (Fig. 1). Digestion with the restriction endonucleases *Sau*3A, *Hinf*I, and *Pst*I gene-rated an identical set of DNA fragments regardless of whether the T4 DNA was derived from pCG57-7 or from the heteroduplex DNA
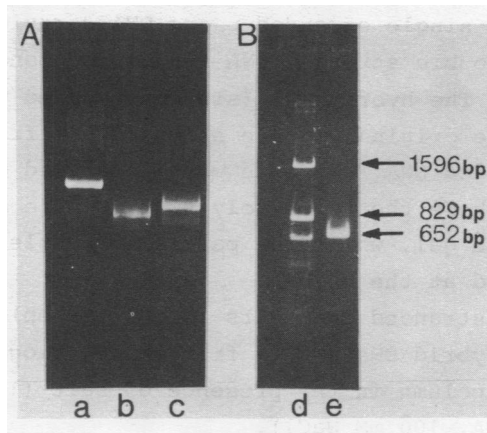
Fig. 1: Heteroduplex formation between single stranded DNAs from fd57-7 and fd57-71 and product analysis after nuclease S1 digestion. Single stranded DNAs were hybridized as described (see METHODS) and analysed on a 1.3 % agarose gel (A). a = hybrid, b = fd57-7, c = fd57-71. The double stranded DNA insert of the hybrid is shown after S1 digestion in panel e of the 2 % agarose gel (B). d = fd RF *Hpa*II digest as sizemarker. Agarose gels were run in E-buffer (0.02 M sodium acetate, 0.04 M Tris base, 0.002 M EDTA adjusted to ph 8.3 with acetic acid).

(Fig. 2).

Separation of 5′ end labelled double stranded DNA fragments.

Double stranded DNA fragments, which were labelled on its 5′ ends for DNA sequencing were hybridized against an excess (molar
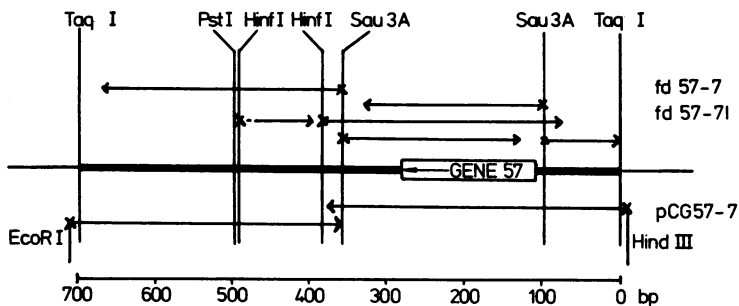
Fig. 2: Restriction map of the cloned T4 DNA *Taq*I E fragment and an outline for DNA sequencing. The arrows indicate length and direction of a sequence. Crosses mark the 5′ labelled end of a fragment. Sequence data derive from fd57-7, fd57-71, and from plasmid pCG57-7 as shown.

ratio 1 : 10) of single stranded phage DNA carrying DNA comple-
mentary to the double stranded DNA fragments. Conditions were as
described above. The hybrid DNA (single stranded phage DNA plus
one strand of the original double stranded DNA fragment(s)) and
the single stranded DNA fragments were separated on a 6 % poly-
acrylamide gel where the relatively small single stranded DNA
migrated into the gel, while the partially single stranded hy-
brid DNA remained at the origin.

   Small single stranded fragments (up to 100 bp) were also sepa-
rated from the hybrid DNA by gel filtration (Biogel A5) in a
pasteur pipette column in the presence of salt (10 mM Tris HCl,
pH 8, 0.1 mM EDTA, 100 mM NaCl).

DNA sequencing.

   DNA was sequenced essentially as described by Maxam and Gil-
bert (7). The sequences were analysed on thin gels (8) and the
gels dried prior to autoradiography (9). Sequence data were pro-
cessed using the computer programs of Osterburg and Sommer (10).

Enzymes and chemicals.

   The restriction endonucleases *Sau*3A, *Hinf*I, *Alu*I, *EcoR*I, and
*Hind*III were purified as described by Roberts (11). Restriction
endonuclease *Pst*I and T4 polynucleotid kinase were purchased
from Boehringer Mannheim GmbH, restriction endonuclease *Hae*II
from Biolabs and S1 nuclease from Sigma Chemie. $\gamma-^{32}P$ ATP was
prepared by E. Beck (12).

RESULTS AND DISCUSSION

DNA sequence and possible proteins.

   The strategy for sequencing is outlined in Fig. 2. The
complete DNA sequence has been established from both strands.
Data derived from double stranded plasmid DNA and from the he-
teroduplex DNA of the single stranded phage DNA are identical.
The junction sequences of the cloned fragment derive from plas-
mid pCG57-7 only, since DNA fragments, which were labelled at 5'
ends deriving from S1 digestion could not be used for sequencing
because of the heterogeneity of those ends. The nucleotide se-
quence of the cloned T4 DNA fragment and the primary structure
of the polypeptides, which can be deduced from that nucleotide se-
quence, are shown in Fig. 3 and Table 1.

Fig. 3: Nucleotide sequence of the cloned T4 DNA *Taq*I fragment. The DNA strand corresponding to the mRNA is shown in the 5'→3' direction from left to right. Deduced polypeptides are listed in frame R1, R2, and R3. The proposed amino acid sequence for gp57A is underlined. The arrows indicate the two possible initiation points for protein synthesis. Sequence homologies to the consensus sequences of bacterial promoters are boxed in (-10 and -35 position) and the translational signal sequence is underlined and marked SD (Shine-Dalgarno sequence). ***indicates termination codons.

DISCUSSION

From the DNA sequence several polypeptides in the range bet-

Table 1: List of polypeptides derived from the nucleotide sequence of the cloned T4 DNA fragment.

Start		Stop		N (AA)	Mol. Weight
ATG	64	TAA	136	24	2973.30
ATG	100	TAA	136	12	1604.20
ATG	109	TAA	136	9	1158.60
ATG	169	TAG	217	16	1903.90
ATG	232	TAA	274	14	1744.60
ATG	9	TAA	285	92	10448.40
ATG	105	TAA	285	60	6459.40
ATG	123	TAA	285	54	5843.60
ATG	287	TGA	527	80	8732.00
ATG	380	TGA	527	49	5246.40

Polypeptides starting after pos. 506 (*Pst*I site) and polypeptides being smaller than 1000 daltons are not included.

ween 1000 and 10000 daltons can be deduced. All polypeptides
coded by DNA extending beyond nucleotide 506 can be probably
eliminated as possible candidates, since a T4 DNA fragment ter-
minating at this position (*Pst*I site) does still complement pha-
ge T4 ambers in gene 57 (3). The most likely candidates (under-
lined in Fig. 3), that fit best the available genetical and bio-
chemical data about gene 57 and gp57A, are two polypeptides of
molecular weights of 5843 and 6459 daltons. Their coding sequen-
ces start at position 105 and 123, respectively, but both ter-
minate at position 285 using the same reading frame. Their mole-
cular weights agree well with the molecular weight of the single
polypeptide (about 6000 daltons), that is synthesized *in vivo* in
E. coli under direction of the cloned T4 DNA fragment (3). The
direction of protein synthesis for the supposed gp57A is in
accordance with the direction of transcription of that region in
T4 infected *E. coli* (13). The amino acid composition of the sup-
posed protein (Fig. 3) is also consistent with the electrophore-
tical behaviour of gp 57A, which migrates in a nondenaturing po-
lyacrylamide gel towards the anode at a pH of 8.3, indicating
that gp57A is not a basic protein.

    All these data support the notion that gp57A is one of the
two suggested proteins. It is also conceivable that both pro-
teins are synthesized because of imprecise chain initiation
as shown for gene 0.3 protein of phage T7 (14) or for gene 2
protein of phage fd (15). Final proof for the identity of gp57A
can only be derived from amino acid analysis of the N-terminus
of that protein. Cloning experiments have shown that gene 57 is
expressed independently of its orientation in the phage fd vec-
tor (3). This suggests that the signal sequences for transcrip-
tion (16) and translation (17, 18) are contained in the cloned
T4 DNA fragment. Possible signal sequences for transcription and
translation of the gene are marked in the T4 DNA sequence (Fig.3)

    Genetical data had provided some hint, that a second protein,
gp57B, might be controlled by   gene 57 (3). The DNA sequence
reveals the possibility of a further protein encoded in the
same reading frame as the proposed gp57A which also terminates
at nucleotide 285. This protein must be initiated outside of the
cloned *Taq*I restriction fragment. Therefore, any amber mutation

in the overlapping part of these proteins should cause loss of two proteins as observed in analysis of T4 amber mutants in gene 57. This second protein is most probably nonessential for tail fiber morphogenesis (3). It has to be initiated in a distance of about 300 bp from the *Taq*I site if it is indeed gp57B. The next known essential gene adjacent to gene 57 is gene 1 (19, 20). Further sequence analysis in that region should clarify, whether there is any relation (e.g. overlap) between gene 1 and gene 57.

The sequence data also explain why amber mutants in gene 57 fail to complement each other (2). Only two nucleotide triplets are present in the proposed reading frame for gp57A (at pos. -195 and 276), which can be converted to an amber codon by mutation of a single base. The triplet at position 276 codes for the antepenultimate amino acid of the supposed gp57A. It is conceivable that gp57A lacking the three C-terminal amino acids may still be functional. If this is true, than is only one potential conditional lethal amber codon present in the coding region for the proposed gp57A. This notion agrees with the observed genetical data (2).

ABBREVIATIONS

wt, wild type; bp, basepairs; SDS, sodium dodecyl sulfate; gp, the polypeptide product of a phage gene.

ACKNOWLEDGEMENTS

REFERENCES

1 Wood, W.B., Bishop, R.J. (1973) in: Virus Research (C.F. Fox and W.S. Robinson, eds.) Academic Press, pp 303 - 326
2 Revel, H.R., Herrmann, R., Bishop, R.J. (1976) Virology 72, 255 - 265
3 Herrmann,R., Wood, W.B. (1981) Molec. gen. Genet. 184, 125-132
4 Herrmann, R., Neugebauer, K., Pirkl, E., Zentgraf, H., Schaller, H. (1980) Molec. gen. Genet. 177, 231 - 242
5 Kupersztoch-Portnoy, Y.M., Lovett, M.A., Helinski, D.R. (1974) Biochemistry 13, 5484 - 5490
6 Shenk, E.T., Rhodes, C., Rigby, P.W., Berg, P. (1975) Proc.

Natl. Acad. Sci. USA 72, 989 - 993
7 Maxam, A.M., Gilbert, W. (1980) Methods in Enzymol. 65, 499 - 560
8 Sanger, F., Coulson, A.R. (1978) FEBS Letters 87, 107 - 110
9 Garoff, H., Ansorge, W. (1981) Analyt. Biochem. submitted
10 Osterburg, G., Sommer, R. (1981) Computer programs in bio-medicine 13, 101 - 109
11 Roberts, R.J., Breitmeyer, J.B., Tabacknik, N.F., Myers, P.A. (1975) J. Mol. Biol. 91, 121 - 123
12 Johnson, R.A., Walseth, T.F. (1979) Adv. Cyclic Nucleotide Res. 10, 135 - 167
13 Fukuda, K., Gossens, L., Abelson, J. (1980) J. Mol. Biol. 137, 213 - 234
14 Dunn, J.J., Buzash-Pollert, E., Studier, F.W. (1978) Proc. Natl. Acad. Sci. USA 75, 2741 - 2745
15 Meyer, T.F., Beyreuther, K., Geider, K. (1980) Molec. gen. Genet. 180, 489 - 494
16 Rosenberg, M., Court, D. (1979) Ann. Rev. Genet. 13, 319-353
17 Shine, J., Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342 - 1346
18 Steitz, J.A., Jakes, K. (1975) Proc. Natl. Acad. Sci. USA 72, 4734 - 4738
19 Wood, W.B., Revel. H.R. (1976) Bacteriol. Reviews 40, 847-868
20 Velten, J. (1981) Ph.D. Thesis, University of California, San Diego