Volume 5 Number 9 September 1978 Nucleic Acids Research

Nucleotide sequence of the O gene and of the origin of replication in bacteriophage lambda DNA

Gerd Scherer+

Institut fur Biologie III der Universitat Freiburg, D-7800 Freiburg i. Br., GFR

Received 11 July 1978

ABSTRACT

The nucleotide sequence of the O gene in bacteriophage lambda DNA is presented. According to two possible initiator codons, the primary structure of the O protein deduced from the DNA sequence consists of 278 or 299 amino acid residues. Structure and function of the O protein - one of the two phage initiator proteins for lambda DNA replication - are discussed in the light of a secondary structure model for the O protein.

The central part of the O gene contains a cluster of symmetrical sequences extending over 160 base pairs. The point mutation of the cis-dominant replication mutant ti12 is located in this region.

INTRODUCTION

On infection of an E.coli host cell, the DNA molecule of bacteriophage lambda (λ), which is linear inside the phage head, becomes circularized and is replicated in two stages : an early, bidirectional circle replication is later followed by an unidirectional rolling circle mode of replication[1]. In contrast to the late phase[1], initiation of the circle replication is confined to a unique region on the λ genome, termed ori (origin of replication). Electron microscopy of replicating λDNA molecules[2] and analysis of deletion prophages incapable of autonomous replication even in the presence of all diffusible elements[3] have located the ori region near 80% on the λ physical map. A set of cis-dominant replication mutants has been isolated, all of which map in this area[4-7]. Recently cloning experiments have shown that an EcoRI fragment of λDNA extending from the immunity region to an EcoRI site at 81.0% of the λ map (fig.1a) contains a functional λ origin[8].

In addition to an intact ori region, initiation of λ circle replication is absolutely dependent on two phage coded functions, the products of genes O and P[4]. Therefore, the λ genome contains the two basic regulatory elements of a replicon[9]: specific initiation factors (the initiator proteins O and P), and a replicator site on the DNA with which the initiators interact (the ori region). Besides the phage proteins O and P, a number of host gene products is essential for λDNA replication (see ref. 1 for review). Finally, local rightward transcription in the ori region seems to be directly required for initiation of circle replication[6,10].

A basic step towards an understanding of the initiation of λDNA replication (in the following, λ(DNA) replication always refers to the early circle replication) in molecular terms is the determination of the primary structure of the regulatory elements involved, i.e. of the initiators (the products of genes O and P) and the replicator (ori). While part of the O gene[11] and part of the ori sequence[12] have already been published, here the entire nucleotide sequence of the O gene and the ori region in phage λDNA is presented. Sequence analysis of a mutant defective in the origin of replication confirms earlier reports[8] that the replicator lies inside the O initiator gene.

MATERIALS

E.coli K12 strains 490A (λdvh93) and 1100 mec$^-$ (λdvh93) were provided by G. Hobom. λdvh93 DNA was isolated as described in ref.13. DNA of phage λti12 (λNam7cI857ti12Sam7) was isolated by heat induction of the lysogen C600 (λNam7cI857ti12Sam7)/λ, a gift of M. Furth. DNA of phages λc26 and λimm21 were kindly provided by G. Hobom.

Restriction enzymes Hind II, Hinf I, Hpa II and Mbo II were prepared by a modification of the procedure of Smith and Wilcox[14]; Hha I and Bgl II, prepared similarly, were gifts of E. Schwarz and R. Grosschedl, respectively. Taq I was isolated according to Bickle et al.[15]. Some BglII and Taq I enzyme was initially provided by V. Pirrotta. EcoRI was obtained from Boehringer.

Alkaline phosphatase from calf intestine (grade I) and T4 polynucleotide kinase were purchased from Boehringer and Biogenics Research Corporation, respectively. DNA polymerase I from E.coli was a gift of L. Loeb.

[γ-^{32}P] ATP at a specific activity of 1000-1500 Ci/mmole was prepared by the procedure given in ref.16, using HCl-free, carrier-free $^{32}P_i$ from New England Nuclear. [α-^{32}P] dATP (250Ci/mmole) was from Amersham. Agfa Gevaert Osray T4 films were used for autoradiography.

METHODS

EcoRI* cleavage. To obtain complete cleavage at EcoRI* sites, up to 400 units of EcoRI/μg DNA were used under the buffer conditions given in ref.18.

Isolation of restriction fragments. The fragment mixtures obtained after restriction enzyme cleavage of λdvh93 DNA were separated on 7.5% or 10% polyacrylamide slab gels (20x30x0.4cm) and the desired fragments were extracted as described[19]. For sequence analysis of λ phage DNA, the DNA was cut with Bgl II and EcoRI enzyme, and the fragments Bgl II-E[17] (nucleotide positions 82-732), Bgl II-G (733-792) and a 354 base pair long Bgl II+EcoRI fragment (793-1146) were isolated from a 7.5% polyacrylamide slab gel.

Terminal labelling. Restriction fragments were labelled at their 5' ends by using polynucleotide kinase and [γ-^{32}P] ATP[16]. 3'terminal labelling of Hinf I-, EcoRI- and Bgl II-fragments was achieved by use of DNA polymerase I and [α-^{32}P] dATP as described[11]. The labelling efficiency in the polymerase reaction was always much better than in the kinase reaction.

DNA sequencing. To separate the labelled ends, the fragments were either cut with an appropriate restriction enzyme, followed by gel electrophoretic separation of the products, or subjected to strand separation as described[16]. The isolated subfragments/single strands were sequenced using the G-, strong A/weak C (A > C) , C-and C+T-cleavages described by Maxam and Gilbert[16]. The cleavage products were processed on 40x20x0.2cm polyacrylamide/7 M urea slab gels, using gel concentrations of 20% or 12% for reading up to 80 or up to 120 nucleotides from

the labelled end, respectively. Before use, gels were aged for one day and pre-electrophoresed at 600 V for at least 6 hours; after loading, the voltage was immediately adjusted to 1000 V.

RESULTS AND DISCUSSION

Sequence determination

As in previous work[11,19], DNA of the λ-derived plasmid λdvh93[13] (fig.1a) was used as a source for restriction fragments. After construction of cleavage maps (G. Scherer and E. Schwarz, unpublished, and ref.13), selected DNA fragments were labelled with ^{32}P at their 5' or 3' ends (fig.1b) and sequenced according to the method of Maxam and Gilbert[16]. Examples of sequence auto-

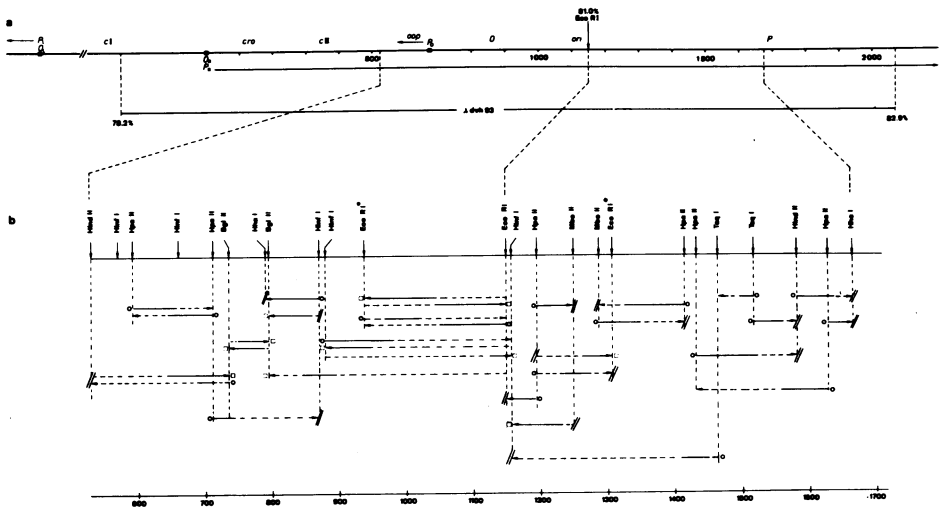


Fig.1a. Physical and genetic map of the immunity and replication region of phage λ. The position of the λ-derived plasmid λdvh93 and of its single EcoRI site at 81.0%[33] is indicated. Transcripts started at promoters P_L, P_R and P_Q are symbolized by horizontal arrows. b. Restriction fragments used for sequence analysis of the Q gene. Clevage sites are indicated by vertical arrows. Horizontal arrows represent terminally labelled single strands pointing from 5' to 3'. Circles and squares symbolize labelled 5' and 3' ends, respectively. After labelling, the fragments were either strand separated (not indicated) or cut by a second restriction enzyme at the points marked //. The solid part of each arrow indicates the unambiguously sequenced part of a fragment. The numbers identify nucleotide positions relative to the P_R mRNA startpoint[11].

radiograms are shown in fig.2, and the final sequence is compiled in fig.3.

As is evident from fig.1b, overlapping fragments were used to get a continuous nucleotide sequence. With the exception of positions 870-877 and 1430-1435, every region was sequenced at

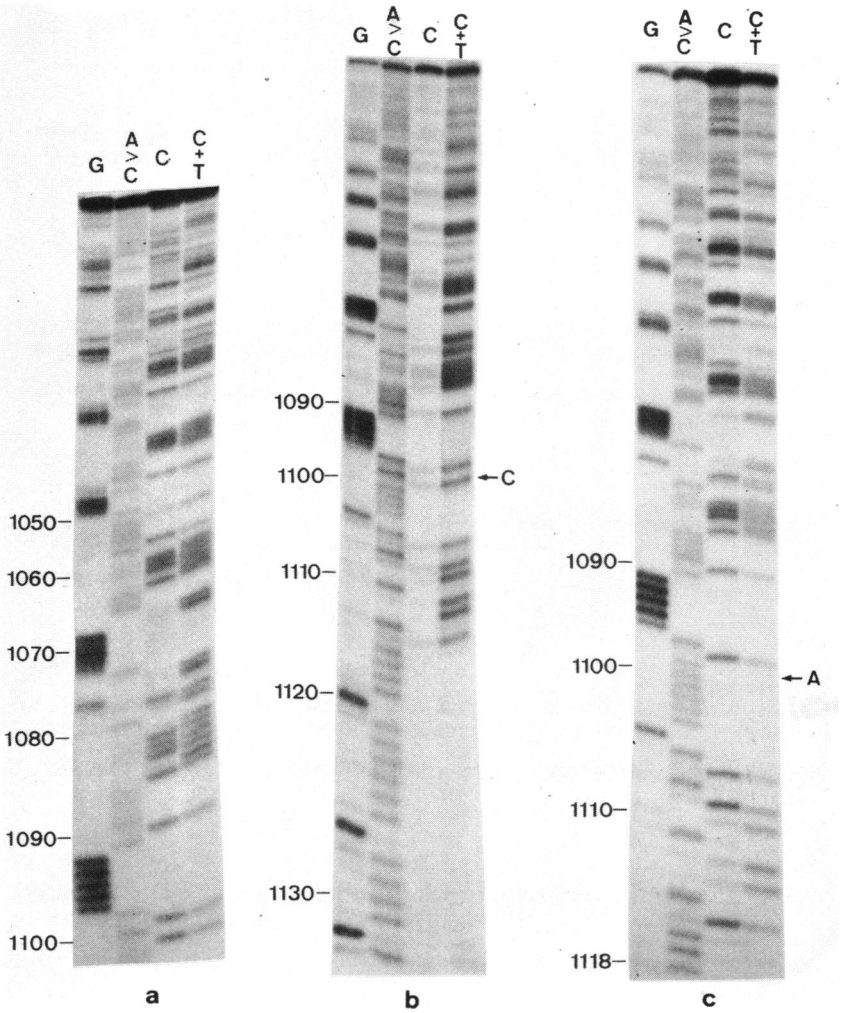
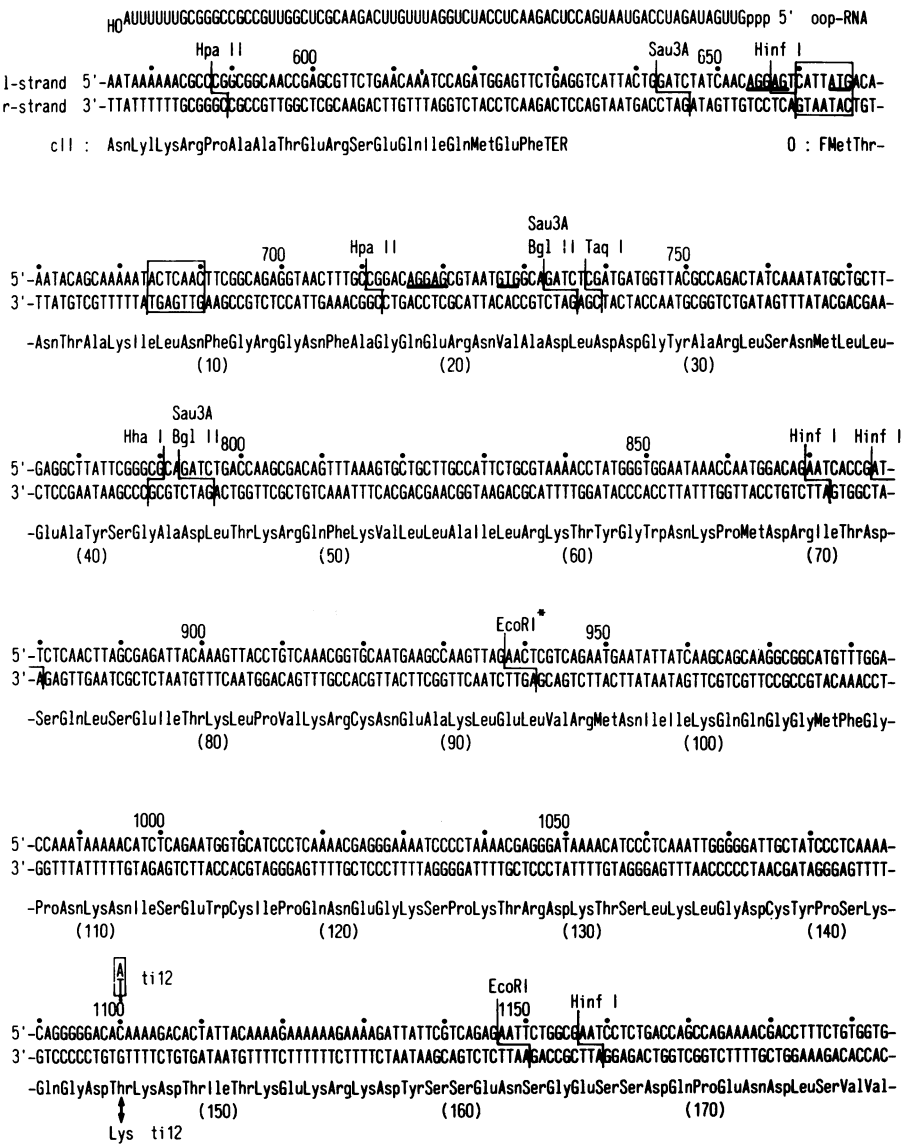


Fig.2. Sequence autoradiograms from the replication region of λDNA. a and b: Patterns obtained from 3' terminally labelled fragment EcoRI* 938-1146 (1-strand positions) derived from λdvh93 DNA. c: sequence patterns from subfragment Hinf I/EcoRI 878-1146 of the 3' terminally labelled fragment Bgl II+EcoRI 793-1146, isolated from λti12 DNA. The ti12 transversion (c) and the corresponding wild type position (b) is marked by an arrow. Gels used contained 12% polyacrylamide/7M urea.

least twice, and about 60% of the sequence was determined from both strands. In addition, various parts of the sequence have also been analysed using bacteriophage DNAs. No sequence differences between phage and plasmid DNA have been observed. The only part of the sequence shown in fig.3, which was not derived from λdvh93 DNA, is the region 733-792.



The recognition sites for the EcoRII restriction enzyme are subject to modification by the C-specific mec$^+$ (dcm$^+$)methylase of E.coli K12[20]. As 5-methylcytosine leads to a gap in the C-

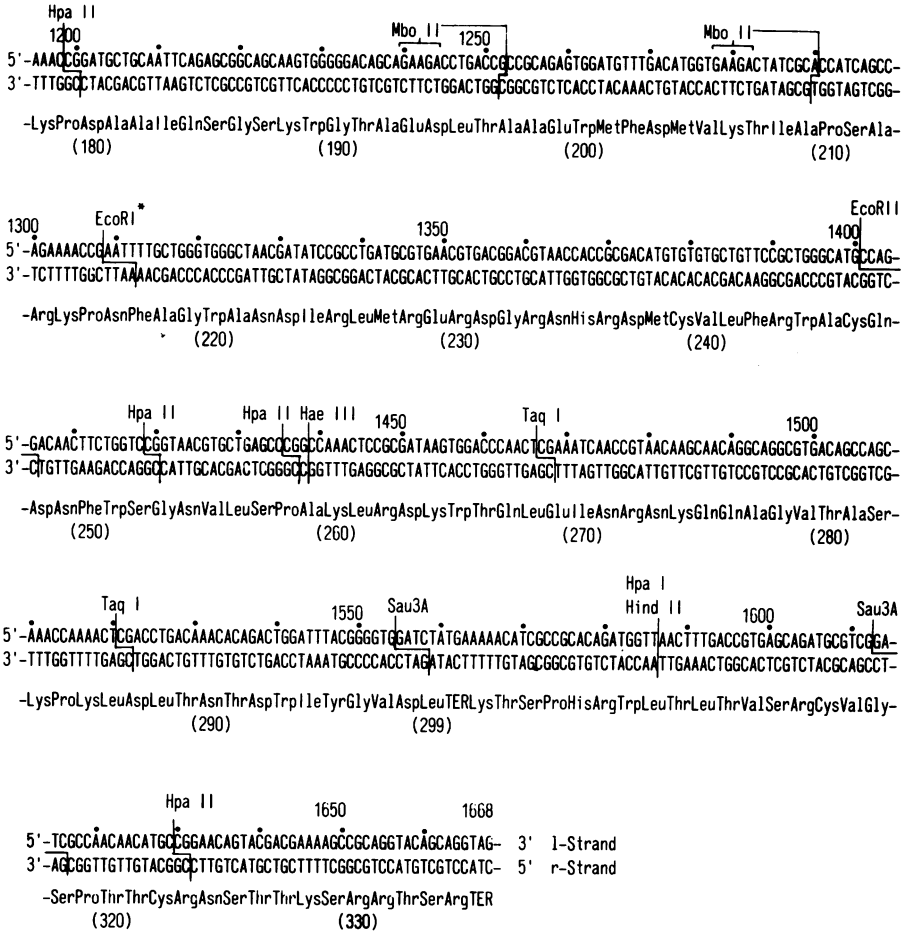


Fig.3. Nucleotide sequence of gene O. The sequence to the left of position 961 has already been published[11]. The C-terminal end of the cII protein sequence and the oop RNA sequence are included[11]; recognition and binding sequences of the oop RNA (p_O) promoter[19] are shown in boxes. Potential initiator codons for gene O and their corresponding ribosomal binding sequences are underlined in the messenger-like l-strand sequence. The amino acid sequence of the O protein and of the read through O'protein is given below the DNA sequence. The transversion of the ori$^-$ mutant ti12 at position 1100 (fig.2) and the exact positions of the restriction sites (fig.1b) are indicated. Nucleotides are numbered relative to the p_R mRNA startpoint (fig.1).

and C+T-lanes of the sequence autoradiograms[21] - which could be overlooked -, the sequence CCTGG at position 1405-1401 (r-strand) in fig.3 was confirmed by sequencing fragment Hpa II 1199-1418 from unmodified λdvh93 DNA, grown in a mec⁻ (dcm⁻) derivative of K12[20]. This is one of three EcoRII sensitive sites in unmodified λdvh93 DNA, the other two lying outside the region sequenced here (G. Hobom, personal communication).

As the enzymes available were not sufficient for complete sequence determination of fragments Hinf I 878-1156 and Hpa II 1199-1418, the isolated fragments were incubated with EcoRI enzyme under conditions of reduced sequence specificity (the EcoRI* activity of Polisky et al.[18]). Of the three sites cleaved (data not shown), only two had a central AATT-sequence : at position 1146 (classical EcoRI site GAATTC) and 1308 (GAATTT). The third EcoRI* cleavage occured at the sequence GAACTC at position 937 (fig.3). Therefore, the target for EcoRI* activity is not simply an AATT-sequence[18], and the loss of specificity under EcoRI* conditions seems not to be confined solely to the outside positions of the EcoRI hexanucleotide sequence.

O gene and O protein

The molecular weight of the O protein as determined by SDS gel electrophoresis is given as 34,500[22] or 37,000[23], respectively, corresponding to about 310-340 amino acid residues. Furthermore, the EcoRI site at 81.0% of the λ physical map (fig.1a) lies inside gene O[8]. Thus the O gene should comprise about 900-1000 nucleotides extending to both sides of the EcoRI site.

By screening the nucleotide sequence shown in fig.3 for termination codons to the left and to the right of the EcoRI site, two potential reading frames for the O protein can be excluded, while the remaining frame is defined by the two in phase nonsense codons at position 607 and 1561. The nonsense codon at position 1561 is an UGA codon, which has been identified as the terminator codon for the O gene[22]. As previously argued[11], there are two possible initiator codons for the O protein: the AUG at position 664 and the GUG at position 727, as both are preceded by sequences complementary to the 3'terminus of ribosomal 16S RNA[24,25] (underlined in fig.3).

The amino acid sequence of the O protein as derived from the nucleotide sequence is included in fig.3. Starting at the AUG codon at position 664, the O protein is predicted to consist of 299 amino acid residues, giving a molecular weight of 33,830 (including the N-terminal methionine), which is in good agreement with the SDS gel estimates of 34,500[22] and 37,000[23]. If translation starts at the GUG codon at position 727, the number of amino acid residues is reduced to 278 and the molecular weight to 31,505.

In an *in vitro* protein synthesizing system using λDNA as template, read through of the UGA terminator codon of the O gene mRNA occurs, producing a protein (termed O') slightly larger than the O protein itself, with a SDS gel estimate of 36,500 (as compared to 34,500 for the O protein)[22]. As can be seen in fig.3, the next in phase terminator codon following the UGA codon at position 1561 is an UAG amber codon at position 1666. This would give a protein 35 amino acid residues or 4,000 daltons larger than the O protein. It is not known whether this read through occurs *in vivo*[22].

29% of the 299 amino acid residues of the O protein are charged residues: 47 residues (19 Arg, 28 Lys) bear a positive and 40 residues (25 Asp, 15 Glu) a negative charge; hence it is a slightly basic protein. In order to get some secondary structure information, the rules of Chou and Fasman[26,27] have been applied to the amino acid sequence. This analysis[28] led to the secondary structure model of the O protein shown in fig.4.

There are 34% helical, 17% ß-sheet and 41% coil residues in this predicted conformation with 41% of the residues participating in ß-turns. A striking aspect of the structure is the clustering of helical and ß-sheet regions in the N- and C-terminal part of the protein, separated by a central region (residues 107-170) consisting almost exclusively of coil and ß-turn residues. (As ß-turns occur mostly at the borders of α- and ß-regions stabilizing neighbouring structural regions[26,27], it is doubtful whether all of the 17 ß-turns predicted for the region 107-170 - which has only three short ß-sheets (fig.4) - will form out). Possibly, the O protein is subdivided into two

structural domains, as is often found in globular proteins: a "N-domain" from residues 1-106 and a "C-domain" from residues 171-299. The secondary structure content in these regions amounts to 49% α-helix and 13% ß-sheet for the N-domain, and 40% α-helix and 19% ß-sheet for the C-domain.

As discussed in ref.12, the O protein appears to be bifunctional, with the N-terminal portion of the protein determining type specificity and presumably recognizing the DNA of the origin region, while the C-terminal part seems to interact with the replication apparatus of the cell, in conjunction with the λP protein. It is tempting to correlate these functions with the N- and C-domains of the O protein predicted here. Accordingly, the "coil"-region 107-170 would simply connect these functional domains, without being of any greater structural significance.

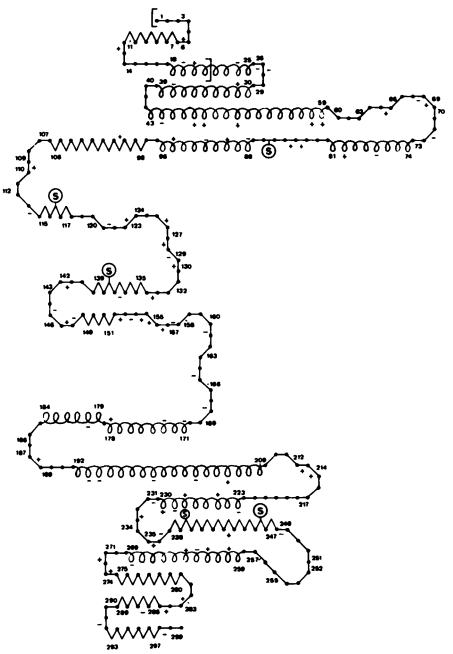


Fig.4. Schematic diagram of predicted secondary structure for the O protein. ℓ, ⩚ and ⟿ represent one α-helix, ß-sheet and coil residue, respectively. ß-bend tetrapeptides are indicated by chain reversals. Charged residues are indicated by + and -, cystein residues by S. N- and C- terminal residues of structural regions are numbered. The first 21 residues are enclosed in brackets, as the exact initiator codon for the O gene is not known (fig.3).

Tests for the functional stability of O and P functions indicate that O is much less stable than P[29]. In the light of the secondary structure model of the O protein, this instability could be explained by an enhanced susceptibility of the protein to protease attack due to the extensive central coil region.

Origin of replication

A straightforward way to identify the λ replicator at the nucleotide level is the sequence analysis of mutants defective in the ori site (ori⁻ mutants). The ori⁻ mutant λti12 forms tiny plaques and displays a 5- to 10-fold defect in autonomous DNA replication[6]. As documented by the autoradiogram shown in fig.2c, λti12 is a point mutant: the C at position 1100 of the wild type sequence (l-strand) is changed to A. The corresponding G to T transversion in the r-strand sequence was also verified (not shown).

As is evident from fig.3, the ti12 mutation lies in the middle of gene O, producing a Thr (ACA-codon) to Lys(AAA-codon) exchange in position 146 of the O protein. This amino acid substitution seems, however, not to impair O function, as λti12 prophages complement superinfecting λO⁻ phages well[6].

In fig.5 the nucleotide sequence of the replicator in phage λ, as defined by the ori⁻ mutants, is shown in detail. The positions of the ti12 transversion (this work) and of the ori⁻ deletions r93, r96 and r99 (from ref.12) are indicated. The ti12 mutation lies inside the r99 deletion, in agreement with genetic data showing recombination between ti12 and r93 as well as r96, but not between ti12 and r99[8]. As with ti12 all three ori⁻ deletions, which remove multiples of 3 base pairs, show an O⁺ phenotype[7,12].

That part of the ori sequence determined independently by Denniston-Thompson et al.[12] is indicated in fig.5. Their sequence deviates from the one presented here at several points (l-strand positions):

- the sequence GA at position 1048 and 1049 is missing;
- they indicate 4 instead of 3 A residues following the C at position 1062;

- in place of the A residues at position 1073 and 1079 they have T and G,respectively;
- between position 1138(T) and 1139(C) there is an additional G in their sequence.

Furthermore, they assign the ti12 transversion (which they have sequenced independently) erroneously to position 1098 (position 1451 in their nomenclature).

Nucleotide sequences between position 1069-1146 have been determined not only from λdvh93 DNA but also from λphage DNA (λti12, λc26)and from both the complementary strands (see also figs.1b and 2). All sequences obtained were in complete agreement. It therefore seems highly unlikely that the deviations between the sequence presented here and in ref.12 are due to strain differences.

The sequences around the ori⁻ mutations are highly structured covering about 160 base pairs of DNA (fig.5). The left half of this region is characterized by an 18 base pair block containing a hyphenated inverted repeat. This block is repeated

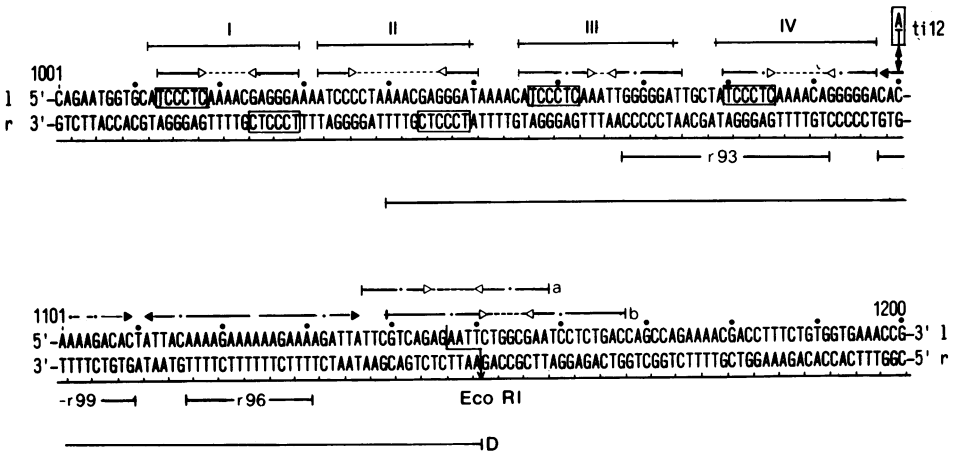


Fig.5. Nucleotide sequence of the ori region in phage λDNA. The position of the ti12 transversion (fig.3) and of the ori deletions r93, r99, r96 (from ref.12) are indicated. The direct repeats I-IV are symbolized by ⟝⟞ ; the common TCCCTC sequence is boxed. Inverted repeats are symbolized by arrows pointing against each other, true palindromes by arrows pointing away from the axis of symmetry. At non-symmetrical positions, the arrows are interrupted by a dot. The reading frame of the O gene is given below the r-strand. D: region sequenced by Denniston-Thompson et al.[12].

four times almost identically (the repeats I-IV, the last two of which were already recognized by Denniston-Thompson et al.[12]). The region following repeat IV (position 1098-1138) has an A/T content of 80% and contains two hyphenated true palindromes (1098-1109 and 1111-1136) located over a 6 nucleotides and 18 nucleotides long pyrimidine tract, respectively. Finally, there are two more, overlapping inverted repeats (a and b) extending over the EcoRI site.

As has been pointed out[12], the A/T rich region mentioned above may contain very important components of the ori region as the strong ori⁻ mutants r99 and r96 lie in this region (fig. 5). The much weaker ori⁻ mutant ti12 sequenced here lies in the same area, producing a C/G to A/T transversion. This base change indicates that the specific sequence and not only the high A/T content has functional importance in this region.

Cloning experiments have shown that an EcoRI fragment of λDNA containing only sequences to the left of the EcoRI site at position 1146 contains a functional λ origin[8]. However, as the replicational activity of cloned λdv fragments containing sequences to both sides of this EcoRI site is 30 times higher as compared to fragments containing sequences only to the left of position 1146[30], and as extensive sequence homology is found around the corresponding EcoRI site in the ori region of the lambdoid phage Ø80 (R. Grosschedl, G. Scherer, G. Hobom, manuscript in preparation), the inverted repeat b - which extends to both sides of the EcoRI site (fig.5) - may be included as a functional element of the λ origin.

The most striking aspect of the λ origin are the inverted repeats I-IV. This structure is reminiscent of the λ operators O_R and O_L, where a more or less symmetrical repressor binding sequence is repeated three times[31]. In both the operators and the ori region, one half of the symmetrical sequence is strongly conserved in the repeats: the half sequence a in the repressor binding sequences[31], the sequence TCCCTC in the repeats I-IV of the ori region (boxed in fig.5). These parallels could point to a multiple binding of a protein factor to this part of the origin, analogous to the multiple binding of repressor molecules to the operators. Not suprisingly, then, the removal of one

of the four repeats in the deletion mutant r93 results only in a partial defect, since this mutant makes minute plaques under favourable conditions[7], as pointed out by Denniston-Thompson et al.[12].

It remains to be seen which role the different parts of the ori region discussed above may play in the initiation of λDNA replication. There is already some evidence that the A/T rich region to the left of the EcoRI site represents a dnaG protein binding site, as this region shows sequence similarity to the origin of complementary strand synthesis in phage G4[32]. Both the λ[1] and the G4[32] initiation of DNA replication is dnaG dependent.

Finally, it should be pointed out that the ori region covers that part of gene O coding for the predicted coil region 107-170 of the O protein (cf. figs. 3,4 and 5). This correlation, together with the fact that in all four ori mutants sequenced the O function is still intact despite considerable changes in the primary structure of the O protein[7,12], could indicate that the DNA region covered by the O gene originally may have contained two separate genes, one to the left and one to the right of the ori region. In this arrangement, the two functions of the later O protein - interaction with the ori DNA through the N-domain, interaction with the P protein through the C-domain - and the nucleotide sequences of the ori region could have evolved separately. Later the two segments of the O gene were combined under inclusion of the ori region, whose DNA sequence had evolved to an optimal recognition site for replication initiation and therefore does not code for any "meaningful" protein function.

ACKNOWLEDGMENTS

I thank Mrs. E. Schiefermayr for her expert technical assistance. The manuscript was improved by the helpful criticisms of H. Kössel, G. Hobom, V. Pirrotta, A. Klein, E. Schwarz and R. Grosschedl. I also thank G. Hobom, V. Pirrotta and M. Furth for gifts of material. Part of this work was supported by a fellowship from the Studientstiftung des deutschen Volkes.

+ Present address : European Molecular Biology Laboratory, Postfach 10.2209, D-6900 HEIDELBERG, FRG.

REFERENCES

1) Skalka, A.M. (1977) Curr.Top.Microbiol.Immunol.78, 201-237.
2) Schnös, M.and Inman, R.B. (1970) J.Mol.Biol.51, 61-73.
3) Stevens, W.F., Adhya, S. and Szybalki, W. (1971) in The Bacteriophage Lambda, Hershey, A.D., Ed., 515-533. Cold Spring Harbor, New York.
4) Eisen, H. et al. (1966) Virology 30, 224-241.
5) Dove, W.F., Hargrove, E., Ohashi, M., Haugli, F. and Guha, A. (1969) Jap.J.Genet.44, Suppl.1, 11-22.
6) Dove, W.F., Inokuchi, H. and Stevens, W.F. (1971) in The Bacteriophage Lambda, Hershey, A.D., Ed., 747-769. Cold Spring Harbor, New York.
7) Rambach, A. (1973) Virology 54, 270-277.
8) Furth, M.E., Blattner, F.R., McLeester, C. and Dove, W.F. (1977) Science 198, 1046-1051.
9) Jacob, F., Brenner, S. and Cuzin, F. (1963) Cold Spring Harbor Symp. Quant. Biol. 28, 329-348.
10) Thomas, R. and Bertani, L.E. (1964) Virology 24, 241-253.
11) Schwarz, E., Scherer, G., Hobom, G. and Kössel, H. (1978) Nature 272, 410-413.
12) Denniston-Thompson, K., Moore, D.D., Kruger, K.E., Furth, M.E. and Blattner, F.R. (1977) Science 198, 1051-1056.
13) Streeck,R.E. and Hobom, G. (1975) Eur.J.Biochem.57, 595-606.
14) Smith, H.O. and Wilcox, K.W. (1970) J.Mol.Biol.51, 379-391.
15) Bickle, T.A., Pirrotta, V. and Imber, R. (1977) Nucleic Acids Res. 4, 2561-2572.
16) Maxam, A. and Gilbert, W. (1977) Proc.Nat. Acad.Sci.USA.74, 560-564.
17) Pirrotta, V. (1976) Nucleic Acids Res.3, 1747-1760.
18) Polisky, B. et al. (1975) Proc.Nat.Acad.Sci. USA. 72, 3310-3314.
19) Scherer, G., Hobom, G. and Kössel, H. (1977) Nature 265, 117-121.
20) Hughes, S.G. and Hattman, S. (1975) J.Mol.Biol.98, 645-647.
21) Ohmori, H., Tomizawa, J. and Maxam, A.M. (1978) Nucleic Acids Res. 5, 1479-1485.
22) Yates, J.L., Gette, W.R., Furth, M.E. and Nomura, M. (1977) Proc.Nat.Acad.Sci. USA. 74, 689-693.
23) Raab, C., Klein, A., Kluding, H., Hirth, P. and Fuchs, E. (1977) FEBS Letters 80, 275-278.
24) Shine, J. and Dalgarno, L. (1974) Proc.Nat.Acad.Sci. USA. 71, 1342-1346.
25) Steitz, J.A. and Yakes, K. (1975) Proc.Nat.Acad.Sci. USA. 72, 4734-4738.
26) Chou, P.J. and Fasman, G.D. (1974) Biochem. 13, 222-245
27) Chou, P.J., Adler, A.J. and Fasman, G.D. (1975) J.Mol.Biol. 96, 29-45.
28) Scherer, G. (1978) Ph.D.Thesis, University of Freiburg.
29) Wyatt, W.M. and Inokuchi, H. (1974) Virology 58, 313-315.

30) Lusky, M. and Hobom, G., submitted to Gene.
31) Pirrotta, V. (1976) Curr.Top.Microbiol.Immunol.74, 21-54.
32) Fiddes, J.C., Barrell, B.G. and Godson, G.N. (1978) Proc. Nat.Acad.Sci. USA.,75, 1081-1085.
33) Thomas, M. and Davis, R.W. (1975) J.Mol.Biol.91, 315-328.