in gene *O*, close to the Eco RI restriction site. Thus, the gene encoding the initiator of the λ replicon overlaps the replicator. Denniston-Thompson *et al.* report the precise localization of the *ori*⁻ mutations, to the level of nucleotide sequence (*22*).

M. E. FURTH
F. R. BLATTNER
C. MCLEESTER
W. F. DOVE

McArdle Laboratory and
Laboratory of Genetics
University of Wisconsin, Madison 53706

References and Notes

1. W. Dove, E. Hargrove, M. Ohashi, F. Haugli, A. Guha, *Jpn. J. Genet.* **44** (Suppl. 1), 11 (1969).
2. W. Dove, H. Inokuchi, W. Stevens, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 747.
3. A. Rambach and P. Brachet, *C. R. Acad. Sci.* **266**, 926 (1971).
4. A. Rambach, *Virology* **54**, 270 (1973). We thank Alain Rambach and Philippe Brachet for the gift of their *ori*⁻ isolates.
5. S. Hilliker and D. Botstein, *J. Mol. Biol.* **106**, 537 (1976); J. Robert *et al.*, in *RNA Polymerase*, R. Losick and M. Chamberlin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1976), p. 707.
6. W. Dove, *Annu. Rev. Genet.* **2**, 305 (1968); J. Szpirer and P. Brachet, *Mol. Gen. Genet.* **108**, 78 (1970); J. Szpirer, *ibid.*, **114**, 297 (1972).
7. M. Furth, C. McLeester, W. Dove, in preparation.
8. R. Thomas and L. Bertani, *Virology* **24**, 241 (1964).
9. H. Inokuchi, W. Dove, D. Freifelder, *J. Mol. Biol.* **74**, 721 (1973).
10. H. Nijkamp, W. Szybalski, M. Ohashi, W. Dove, *Mol. Gen. Genet.* **114**, 80 (1971); M. Ohashi and W. Dove, *Virology* **72**, 299 (1976).
11. F. Blattner, B. Williams, A. Blechl, K. Denniston-Thompson, H. Faber, L. Furlong, D. Grunwald, D. Kiefer, D. Moore, J. Schumm, E. Sheldon, O. Smithies, *Science* **196**, 161 (1977).
12. D. D. Moore, K. Denniston-Thompson, M. E. Furth, B. G. Williams, F. R. Blattner, *Science* **198**, 1041 (1977).
13. The replication region of Charon 3 includes gene *P* of λ (*6, 11*) and therefore is unable to function in a *gro*P host, while that of φ80 or P22 is able to function [π phenotype, described by C. Georgopoulos and I. Herskowitz, in *The Bacteriophage Lambda*, A. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 553]. Recombinants carrying the late region of a Charon 3 derivative (including the inserted fragment) and the replication region of φ80 or P22 can be selected by plating the progeny of a cross of the Charon 3 derivative by either φ80 *14*am8 (*14*) or λ*h*80 *imm*434 *rep*P22 *12*amN14 (*12*) on a *gro*PA15 derivative of C600 φ80^R *su*II. The *am* alleles of the φ80 and P22 replication systems facilitate mapping and functional tests.
14. K. Sato, Y. Nishimune, M. Sato, R. Numick, A. Matsushiro, H. Inokuchi, H. Ozeki, *Virology* **34**, 637 (1968).
15. R. Thomas and S. Mousset, *J. Mol. Biol.* **47**, 179 (1970).
16. W. Arber, in *The Molecular Biology of Viruses*, L. Crawford and M. Stocker, Eds. (Cambridge Univ. Press, Cambridge, 1968), p. 295.
17. λ*ri*$^+$ recombinants are obtained in crosses of λ*ri*^{c}5b or λ*ri*^{c}D with a derivative of Ch3S45 but not with a derivative of Ch3λ2. [M. E. Furth, in preparation].
18. M. Castellazzi, P. Brachet, H. Eisen, *Mol. Gen. Genet.* **117**, 211 (1972).
19. *O*am905 was isolated by P. Aoothman [thesis, University of Oregon (1976)] and was identified as an *O* mutant on the basis of complementation tests. We were concerned that this mutation might lie in a gene upstream to *O* and that the failure to express *O* function might be due to nonsense polarity. However, we find that *O*am905 fails to complement with other *O*⁻ mutants even in a host carrying the strong polarity suppressor *psu*104 (*28*) (unpublished data of M.E.F. and C.M.). Thus, *O*am905 must lie within gene *O*. We thank P. Toothman, M. Jones, and I. Herskowitz for their gifts of the *O*am905 mutant and of the MJ series of prophage deletions. We thank N. Franklin for sending us the *psu*104 *E. coli* strain. *O*am1005 was isolated in this laboratory among a set of nitrosoguanidine-induced amber mutants of λ (M. E., Furth, unpublished).
20. The deleted prophages (λ*N*am7*N*am53 *c*Its857 Δ) were introduced by P1 transduction into strain 594 *gal*K⁻T⁻ *trp* : : *Tn*5 *su*° (a gift of D. Berg), by selection of *gal*$^+$ and screening for *iomm*λ at 30°C. The amber suppressor *su*III was then introduced by cotransduction with *trp*$^+$ by phage P1 grown on Ymel.
21. Strictly speaking, the failure of *ori*⁻ mutants to recombine with deleted prophages F3 and MJ17 implies that at least one component of each of the *ori* defects lies to the right of the *O*am905 site. The *ori*⁻ mutants all recombine efficiently with *O*am905 (data not shown). It seems highly unlikely that each of these spontaneous *ori*⁻ mutants could involve multiple defects spanning *O*am905. Although we have assumed that the *ori*⁻ mutations directly inactivate the λ replicator, we cannot completely exclude the formal possibility that they prevent expression of an unknown *cis*-acting replication gene (*2*).
22. K. Denniston-Thompson, D. D. Moore, K. E. Kruger, M. E. Furth, F. R. Blattner, *Science* **198**, 1051 (1977).
23. J. Yates and M. Furth, unpublished observations, using the method of J. Yates, W. Gette, M. Furth, M. Nomura, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 689 (1977).
24. In agreement with Rambach (*4*), we find that the class A mutants, *r*95, *r*96, and *r*99, fail to revert frequently (< 10^{-8} at 30°C) and fail to give minute recombinants with λ*imm*21 *O*⁻ *P*⁻. Class B mutants, including *r*93, *r*97, and *r*98, both revert and give minute recombinants with λ*imm*21 *O*⁻ *P*⁻. The higher reversion rate of class B mutants is probably due to second-site mutations.
25. I. Tessman, *Genetics* **51**, 63 (1965) observed that for markers separated by extremely small intervals, recombination frequency may be much lower than expected on the basis of recombination frequencies observed for larger intervals. C. Thomas, *Prog. Nucleic Acid Res. Mol. Biol.* **5**, 315 (1966) presented theoretical arguments that the minimum size of an effective recombinational pairing region should be about 12 to 15 base pairs (bp). Our data are consistent with this model. The apparent recombination frequency observed here for two deletions separated by 24 base pairs is of the order of magnitude expected for general recombination in λ, while the frequencies observed for two deletions separated by six base pairs are two to three orders of magnitude lower than expected.
26. S. Benzer, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 410 (1961).
27. L. Barnett, S. Brenner, F. Crick, R. Shulman, R. Watts-Tobin, *Phil. Trans. R. Soc. London Ser. B*, **252**, 487 (1967).
28. L. Korn and C. Yanofsky, *J. Mol. Biol.* **103**, 395 (1976).
29. This is paper No. 2171 from the Laboratory of Genetics. Supported by grants from the National Cancer Institute to McArdle Laboratory (W.F.D.) and from the National Institute of General Medical Sciences to F.R.B. M.E.F. acknowledges support of a predoctoral fellowship from the National Science Foundation, and a training grant to McArdle Laboratory from the National Cancer Institute. We thank D. Moore and K. Denniston-Thompson for stimulating discussions and for critical reading of the manuscript. We thank D. Berg, P. Brachet, N. Franklin, I. Herskowitz, M. Jones, A. Rambach, and P. Toothman for gifts of bacterial and phage strains. This work was done under NIH guidelines, which call for EK1, P1.

19 July 1977; revised 20 September 1977

Physical Structure of the Replication Origin of Bacteriophage Lambda

Abstract. *The nucleotide sequence of part of the replication region of wild-type bacteriophage lambda and of four mutants defective in the origin of DNA replication* (ori⁻) *has been determined. Three of the* ori⁻ *mutations are small deletions, and one is a transversion. The sequence of the origin region, defined by these mutations, contains a number of unusual features.*

Moore *et al.* (*1*) and Furth *et al.* (*2*) have described the construction and functional analysis of chimeric phages carrying restriction endonuclease fragments from the replication control region of the bacteriophage lambda (λ) genome. Fragments extending from the immunity control (*imm*) region to the Eco RI site within gene *O* (*imm-ori*λ fragments) have been shown to carry a fully functional origin of replication, displaying all the known requirements of the normal λ replicator. All known mutations (*ori*⁻) that inactivate the λ replication origin, without affecting expression of the structural genes (*O* and *P*) essential for replication, lie in a small portion of this fragment (*2*). We now describe a structural analysis of the λ replication region, including a determination of the changes in nucleotide sequence by which four *ori*⁻ mutations inactivate the replicator.

The first step in this analysis was to determine the locations of restriction endonuclease cleavage sites within the 2.75 percent λ Eco RI fragment from Ch3λ2*imm*434Δ7c (or an identical fragment from phage KK1, Fig. 1; *3*). This short fragment contains *imm*434 genes on the left and λ genes on the right, as shown in Fig. 1. A similar Eco RI fragment from the *imm*λ phage KK2 (Fig. 1) is about 3.7 percent λ in length and contains only λ genes. The two types of fragment were, respectively, designated *imm*434-*ori*λ and *imm*λ-*ori*λ. Many restriction endonuclease cleavage sites (*4*) were located by comparing digestion patterns of the two *imm-ori*λ fragments.

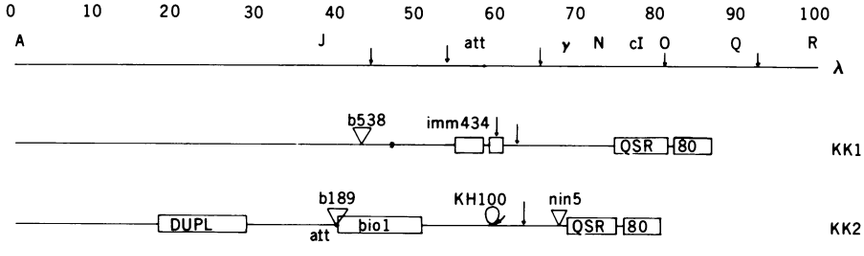
The fragments were purified by differential centrifugation (*1*) and were labeled at the 5′ termini with ^{32}P (*5, 6*). The fragments were then redigested with a second restriction enzyme. The subfragments were separated by electrophoresis on polyacrylamide gels (*7*) and visualized both by staining and by radioautography. Since the right ends of the two *imm-ori*λ fragments are identical, any subfragments with differing

mobilities must come from the left. The terminal subfragments, identified by their ^{32}P label, and the internal subfragments, identified only by staining, were thereby assigned to the left or right ends of the *imm*434-*ori*λ and *imm*λ-*ori*λ fragments. Finally, any remaining ambiguities were resolved by comparison of the sizes of double digestion products with the sizes of subfragments produced by digestion with each enzyme alone.

An example of this approach to mapping restriction sites is the analysis of the Hpa II digestion patterns shown in Fig. 2. This enzyme produced five subfragments of 517, 415, 215, 123, and 95 base pairs (bp) from the *imm*434-*ori*λ fragment. Hpa II digestion of the *imm*λ-*ori*λ fragment also produced 123- and 517-bp subfragments, but a single 1270-bp subfragment appeared instead of the 415-, 215-, and 95-bp species. In each case the 517-bp subfragment was labeled by ^{32}P and was therefore assigned to the right end. The 123-bp subfragment must occupy a position second from the right, since it was the only other subfragment common to both digests. The 215-bp subfragment of *imm*434-*ori*λ was also end-labeled and was therefore assigned to the left end of the map. As was expected, the 1270-bp subfragment was end-labeled in the *imm*λ-*ori*λ DNA. From this experiment, it was possible to determine the Hpa II subfragment order for the *imm*434-*ori*λ fragment: 215-(95, 415)-123-517 (Fig. 3). The Hpa II subfragment order 95-415 was deduced by comparison of the sizes of products of double digestion by Hpa II and a second enzyme (Alu I, Hae II, Mbo I, or Taq I) with the sizes of subfragments produced by single digestions.

The sites of cleavage of the *imm*434-*ori*λ fragment by ten different restriction endonucleases are shown in Fig. 3. The subfragments produced by each of these enzymes are listed in Table 1. We also took advantage of DNA sequence information as it became available, to confirm and refine the restriction map (*8–14*).

We adopted an arbitrary scale for numbering nucleotide pairs by designating the 3′ terminal nucleotide of the *l* strand of *imm*434-*ori*λ fragment (that is, the site of Eco RI cleavage in gene *O*) as 1500 (see Figs. 3 and 5). This placed the



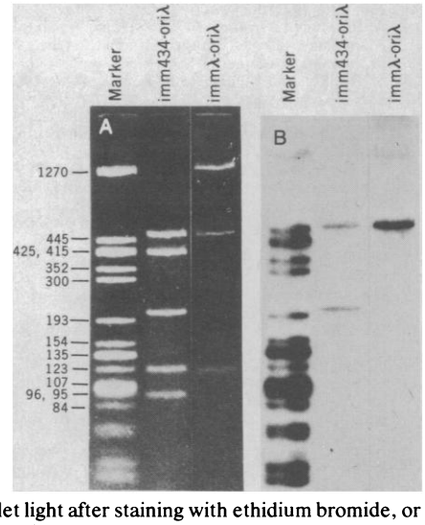


Fig. 1 (above). Physical maps of bacteriophages λ, KK1, and KK2 are shown beneath the 0 to 100 percent λ scale. 1 percent corresponds to about 494 base pairs in duplex DNA. KK1 and KK2 serve as sources of the 2.75 percent *imm*434-*ori*λ and the 3.7 percent *imm*λ-*ori*λ Eco RI fragments, respectively. Downward arrows (↓) represent sites of cleavage by Eco RI. Deletions are shown by inverted triangles. Heterologous substitutions of *E. coli*, 434, or φ80 DNA are represented by boxes. KH100 has an IS5 insertion in gene *c*I that contains an Eco RI restriction site. DUPL is a duplication of λ genes in the left arm. The construction of KK1 and KK2 is described (*3*). Fig. 2 (right). The *imm*434-*ori*λ and *imm*λ-*ori*λ Eco RI fragments from phages KK1 and KK2, respectively, were purified by differential centrifugation (*1*). These fragments were labeled with ^{32}P at the 5′ termini (*5*), digested with Hpa II, and electrophoresed in 10 percent polyacrylamide (*7*). Digestion products were visualized by (A) photography under ultraviolet light after staining with ethidium bromide, or (B) autoradiography. Hpa II fragments of pKDT2, labeled at the 5′ termini with ^{32}P, served as size markers.

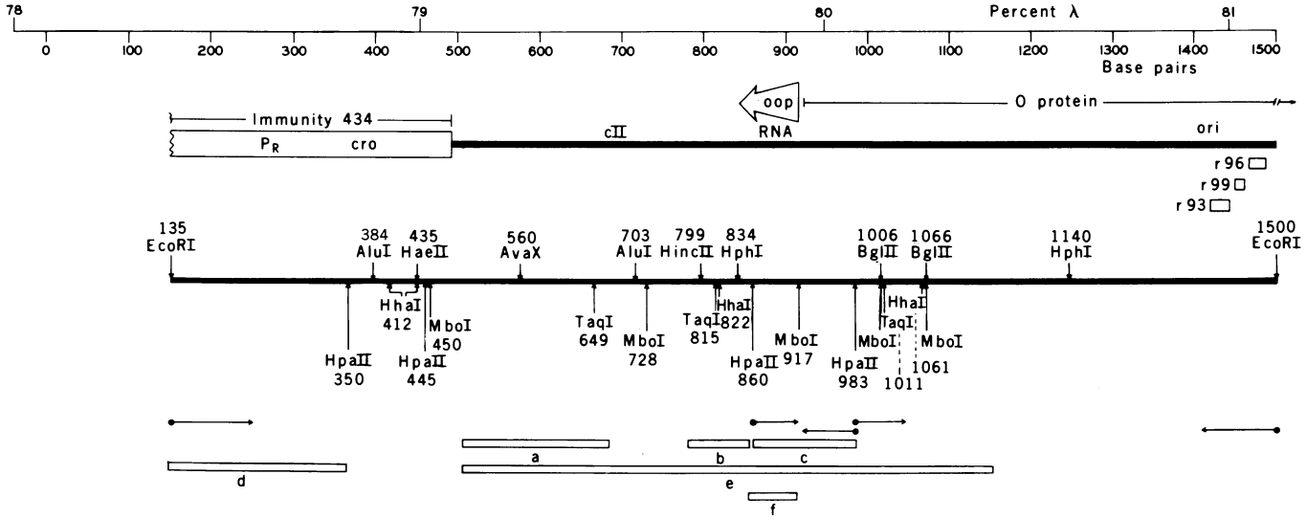


Fig. 3. Map of restriction endonuclease cleavage sites in the *imm*434-*ori*λ fragment, approximately aligned with the genetic map. The percent λ scale identifies nucleotide positions with respect to the rightmost Eco RI cleavage site, arbitrarily defined as 1500 (on the *l* strand). Exact positions assigned to restriction enzyme cleavage sites are given. (Hae II and Bgl II cleavage sites are subsets of Hha I and Mbo I sites, respectively.) Horizontal arrows indicate areas for which we have determined the nucleotide sequence (*6*). Boxes indicate sequences determined by others: (a) M. Rosenberg (*9*); (b) Kleid *et al.* (*10*) and Rosenberg *et al.* (*11*); (c) Scherer *et al.* (12); (d) V. Pirotta (*13*); (e) E. Schwarz, G. Scherer, G. Hobom, and H. Kössel (*14*); and (f) J. Dahlberg and F. Blattner (*8*).

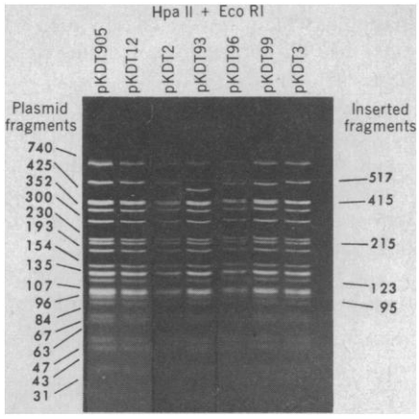


Fig. 4. Polyacrylamide gel electrophoresis of Hpa II and Eco RI double digests of chimeric plasmids. pKDT2, pKDT3, pKDT12, pKDT93, pKDT96, pKDT99, and pKDT905 DNA's (*1*) were digested to completion with an Hpa II–Eco RI mixture. DNA fragments were separated by electrophoresis through a 10 percent polyacrylamide slab gel (*7*).

left end of this fragment approximately at position 135.

The restriction enzyme activity Ava X was useful for relating the restriction map to previous electron micrographic and genetic maps of this region. We discovered this activity as an uncharacterized minor component of a preparation of Ava I. While Ava I does not cut the *imm*434-*ori*λ fragment, Ava X cuts at approximately 80 bp to the right of the *imm*434 substitution, a region for which the sequence has been determined (*9*) (Fig. 3). This allowed us to relate our arbitrary scale of the *imm-ori*λ region to the standard λ physical map, on which the *imm*434-λ junction is a widely used fiducial mark of 79.1 percent (*15*).

Furth *et al.* (*2*) showed that seven independent *ori*$^-$ mutations lie on the *imm-ori*λ Eco RI fragment. We report here that three of these mutations are small deletions located near the right end of this fragment. Figure 4 shows Eco RI plus Hpa II digestion patterns of plasmid DNA's containing the *imm*434-*ori*λ fragment from each of four *ori*$^-$ mutants. Three of them (*r*93, *r*96, and *r*99) have a perceptibly shorter subfragment in place of the 517-bp subfragment of wild-type. This is the rightmost Hpa II subfragment. A similar shift was observed in the mobility of the rightmost Hph I subfragment (360-bp in wild-type; data not shown). One of the *ori*$^-$ mutants (*ti*12), two *O* amber mutants and several wild-type isolates from different sources did not show detectable alterations in mobility of the rightmost Hpa II subfragment.

Plasmid DNA carrying a wild-type or mutant λ *ori* region was digested with Eco RI (*1*) and labeled at the 5′ terminals by phosphorylation with ^{32}P (*5*, *6*). The DNA was then digested with Hpa II to produce subfragments labeled at one (the Eco RI) end. These were separated on 5 percent polyacrylamide gels. The *ori*-containing subfragment was eluted, and the nucleotide sequence was determined by the method of Maxam and Gilbert (*6*). Sequencing gels for the wild-type *ori* region and for two of the *ori*$^-$ mutants are shown in Fig. 5. The sequence of 110 bp to the left of the Eco RI site in *O* was deduced from 15 experiments of this type, all of which gave consistent results (Fig. 6). The results have not yet been confirmed by sequencing the complementary stand or by analysis of the O protein. Thus, unanticipated anomalies in the DNA chemistry could be a source of error.

The sequence of DNA from each of the *ori*$^-$ mutants *r*93, *r*96, and *r*99 shows a simple deletion near the Eco RI site, of 24, 15, and 12 bp, respectively (Fig. 6). The extents of the deletions agree with the changes in mobilities of the *ori*$^-$ DNA fragments (Fig. 4). The DNA sequence of the *ori*$^-$ mutant *ti*12 shows a C · G to A · T change at position 1451. This transversion has been observed independently by Scherer *et al.* (*14a*). No other differences in sequence were observed between the *ori*$^-$ mutants and wild-type λ.

We find no protein initiator codons in any phase between any of the three *ori*$^-$ mutations and the Eco RI site. Because this site lies within gene *O*, this provides strong evidence that these *ori*$^-$ mutations also lie within *O*, in agreement with the genetic mapping presented by Furth *et al.* (*2*).

The sequence of the λ *ori* region (Fig. 6) has a number of unusual features. For example, there is a stretch of 48 nucleotides on the *r* strand of which 81 percent are pyrimidines. This region contains an unbroken tract of 18 pyrimidines, which probably is the longest in λ (*16*). The sequence also has many runs of identical nucleotides. There are seven instances of at least four T's (thymine) in a row, and two runs of five C's (cytosine), on the *r* strand. As is shown in Fig. 6 (arrows), there is a long hyphenated repeat which also contains inverted repeats. This portion of the sequence can be represented as either a single interrupted hairpin loop with 13 paired bases separated by an internal 4-bp mismatch or as two hairpin loops with a total of 18 paired bases (Fig. 7).

In this group of three reports (*1*, *2*) we have identified a region of λ DNA which has two functions: (i) it serves as the λ replicator and (ii) it encodes part of the structural gene *O*. It is clear that the DNA deleted by each of the three *ori*$^-$ mutations *r*93, *r*99, and *r*96 contains a nucleotide sequence essential for *ori* but dispensable for *O*, since the mutants were chosen to retain *O* function (*17*, *18*). The minimum contiguous segment of DNA which includes at least one base pair affected by each of these *ori*$^-$ deletions is 26 bp long. This sets a lower limit to the size of the λ replicator. An upper limit is set on the right by the Eco RI cleavage site (*2*) but is not yet determined on the left.

We do not know which aspect or aspects of origin function is involved in each mutant defect—binding of an initiator, entry of accessory factors, or polymerization of nucleotides. Nevertheless, it is interesting to consider the relation of the strength of the *ori*$^-$ phenotype to the structural features affected in each case.

The λ*r*96 and λ*r*99 mutants are strongly *ori*$^-$ (*2*, *17*), and thus the pyrimidine tracts removed by these deletions may well contain very important components of the origin region. This provides an occasion to recycle the speculations about dPy · dPu (deoxypyrimidine · deoxypu-

Table 1. Sizes of restriction endonuclease subfragments of *imm*434-*ori*λ fragment. Subfragments produced by digestion of purified *imm*434-*ori*λ fragment by each of ten restriction endonucleases were subjected to electrophoresis in 10 percent polyacrylamide gels containing 25 percent glycerol (*7*). Sizes were determined by comparison with products of Hpa II plus Eco RI digestion of plasmid pKDT2, which were calibrated by comparison with a Hae III digest of φX174 (*24*). Numbers not in parentheses are subfragment sizes determined by mobility on gels. Fragments too large to measure accurately are designated NM. Numbers in parentheses are our best estimates of subfragment sizes; these estimates are based both on many single and multiple digestion experiments and on DNA sequence information. Subfragments are listed in their map order (left to right).

Enzyme	Subfragment (bp) sizes
Alu I	(249 319 797)
	250 320 NM
Ava X	(425 940)
	415 NM
Hinc II	(664 701)
	570 720
Hha I	(277 23 387 239 439)
	285 21 385 240 450
Hae II	(300 1200)
	320 NM
Hpa II	(215 95 415 123 517)
	215 95 415 125 520
Hph I	(699 306 360)
	NM 305 360
Mbo I	(315 278 187 89 60 434)
	315 275 190 95 65 440
Bgl II	(871 60 434)
	NM 65 470
Taq I	(514 166 191 489)
	540 165 195 490

rine) clusters discussed by Szybalski (*19*).

λ*r*93 appears to be a slightly less defective mutant since it can make minute plaques under favorable conditions (*2*, *17*). This mutation removes one copy of the partial duplication (it may have arisen by recombinational elimination of the duplication) as well as part of the largest inverted repeat. The residual *ori* function of λ*r*93 could lie either in the remaining copy of the duplicated sequence or in a truncated hairpin loop.

The λ*ti*12 is the least defective *ori*$^-$ mutant we have studied, since it forms tiny plaques under normal plating conditions. More direct measurements show that the *ti*12 mutation impairs λ DNA replication by about tenfold (*18*, *20*). Sequence analysis indicates that the *ti*12 mutation is a transversion at position 1451 (Fig. 6). This agrees with genetic mapping, which shows that *ti*12 lies very close to *r*99, between the flanking deletions *r*93 and *r*96 (*2*).

Kleckner (*21*) reported that, in the absence of *N* function, λ*O*am29 exhibits a *cis*-dominant defect in replication. The *O*am29 mutation maps near the *ori*$^-$ mutations (*2*, *17*, *18*) and so may simultaneously affect *O* and *ori*.

The fact that the strongest *ori*$^-$ mutations studied are deletions is significant,

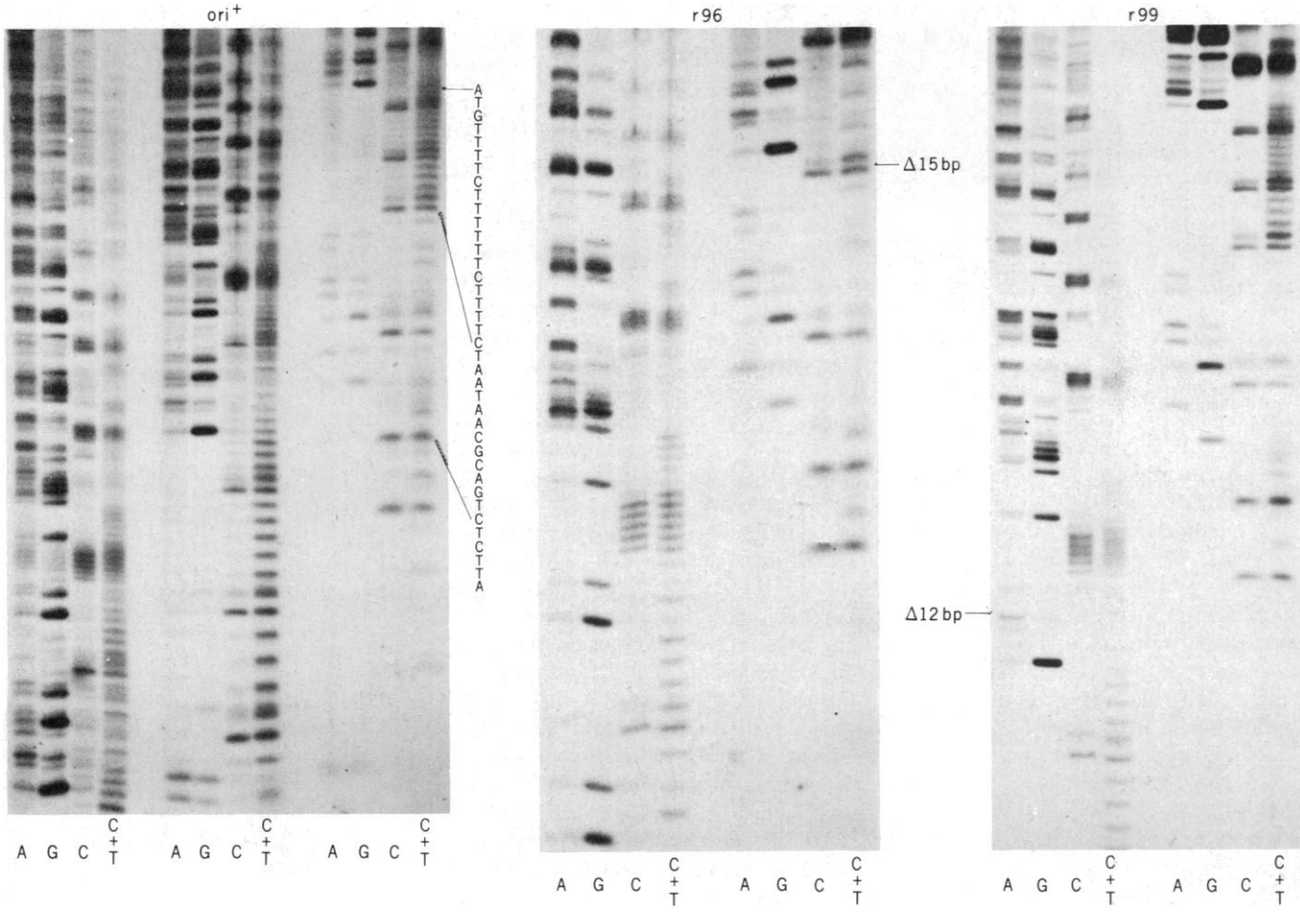
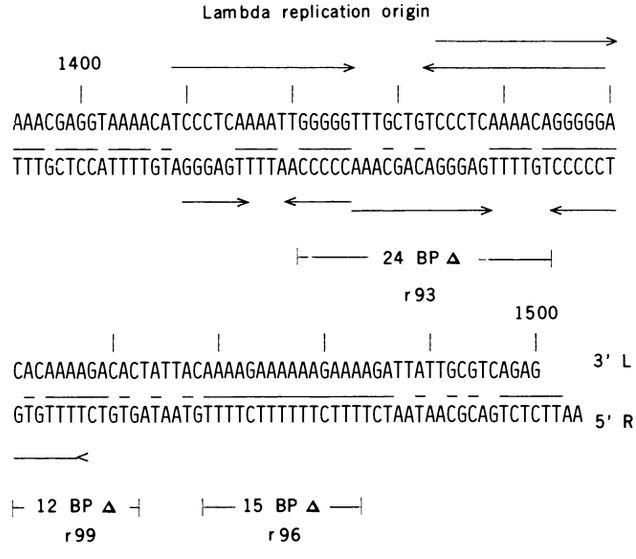


Fig. 5 (above). Autoradiograms of sequence gels of wild-type and two mutant λ *ori* regions. Plasmids containing the *imm*434-*ori*λ fragment (pKDT3, pKDT96, pKDT99) (*1*) were digested with Eco RI. The resulting fragments were end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and redigested with Hpa II. From each strain the Hpa II–Eco RI subfragment containing the *ori* region (517 base pairs in wild-type), labeled only at the Eco RI end, was purified and subjected to the four base-specific chemical degradations described by Maxam and Gilbert (*6*). The products were divided into portions which were resolved by electrophoresis on 20 percent polyacrylamide gels containing 7*M* urea for varying lengths of time (*ori*$^+$ for 48, 30, and 12 hours; *r*96 and *r*99 for 30 and 12 hours). The gels were frozen and the end-labeled DNA degradation products were visualized by autoradiography. Fig. 6 (right). Nucleotide sequence of the replication origin region of bacteriophage λ. The sequence of the region of the *r* strand immediately to the left of the Eco RI cleavage site in gene *O* (position 1504) was determined by the method of Maxam and Gilbert (*6*). Nucleotides deleted in the *ori*$^-$ mutants *r*93, *r*96, and *r*99 are indicated. The *ori*$^-$ mutant *ti*12 shows a single C · G to A · T transversion at position 1451. The sequences determined for the *ori*$^+$ parent of the *ori*$^-$ mutants (cloned in pKDT3) and for another *ori*$^+$ strain (cloned in pKDT2) were identical. Arrows in the same direction indicate a partial duplication (15 of 27 base pairs); arrows in opposite directions indicate inverted repeats (Fig. 7). Lines between strands indicate the positions of pyrimidines on the *r* strand. *A*, adenine; *G*, guanine; *T*, thymine; *C*, cytosine.



since deletions were not explicitly selected (*17, 18*). This may mean that no base pair in the origin region is sufficiently important for a point mutation to produce a tight *ori*$^-$ phenotype.

An interesting question is whether the bidirectionality of λ replication results from a bipartite substructure of the origin. We do not see an extended symmetrical structure within the region delineated by the *ori*$^-$ mutations, but additional sequence analysis might reveal such structure. Conversely, it seems possible that bidirectional replication emerges from a unidirectional initiation event (*22*).

Insight into *ori* function may emerge when a number of comparable origin regions have been characterized in sufficient detail. At present we can only note a tendency of such regions from diverse types of replicons to contain long dPy · dPu clusters (*23*). The long AT (adenine, thymine)-rich dPy · dPu tract removed by the *r*96 deletion (Fig. 6) may facilitate localized strand separation in the origin region.

Like λ, φX174 has its *ori* site within the structural gene for its initiator protein (*24*). Close linkage of these interacting entities may be of importance in evolution (*25*). We cannot yet assess any physiological significance of such linkage.

We have shown both by genetic and by structural analysis that the *ori* region and gene *O* overlap; thus the sequence shown in Fig. 6 reveals information about the O protein. Because the molecular weight of the O polypeptide is about 35,000 (*26*), this sequence corresponds to about 12 percent of gene *O*.

We are not yet able to determine the reading frame of the *O* gene, but one reading frame is not possible due to a translation terminator codon at position 1402. In either remaining reading frame, the segment of the polypeptide corresponding to the known nucleotide sequence would be very rich in basic amino acids (48 percent in one frame and 27 percent in the other). Each of the *ori*$^-$ deletions removes a multiple of 3 bp, so that no phase shift is created. This result is consistent with the *O*$^+$ phenotype of the mutants. In each of the possible reading frames, the deletion at the DNA level translates into a simple amino acid deletion; there are no substitutions introduced by fusion of codons.

We are amazed to find that mutations selected for retention of *O* function would be deletions. However, some properties of the *O* gene suggest that its product is bifunctional, and hence the region of protein affected by *ori*$^-$ mutations

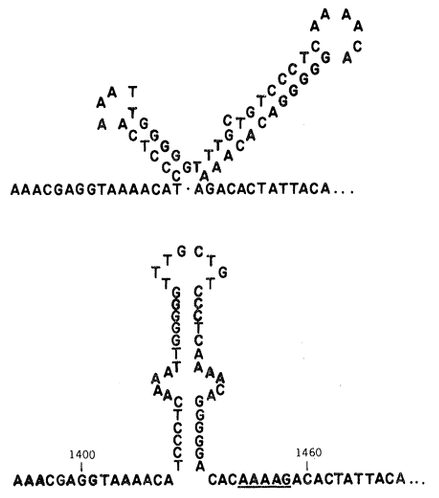


Fig. 7. Structures of potential hairpin loops located between positions 1409 and 1460. The underlined sequence A_4G is similar to the sequence at the junction between the RNA primer and DNA seen in *Col* El DNA synthesis in vitro (*23*).

might serve as a connector. The NH_2-terminal portion of the O protein determines type specificity and presumably recognizes the DNA of the origin region (*27*). The COOH-terminal portion appears to interact with the replication apparatus of the cell, through or in conjunction with the P protein (*28*). Since *ori* is near the junction between the two functional domains of *O* (in the same interval as the Eco RI site) (*2, 27*), the region of the protein encoded by this part of gene *O* may simply connect the two functional parts. Such a passive protein function might relieve coding constraints on the sequence so that it could evolve DNA function as a recognition site for replication.

The role of RNA synthesis in λ DNA replication is not yet understood. It has been suggested that *oop*, a small RNA molecule encoded near the replication region, primes DNA replication (*29*). However, all of the *ori*$^-$ mutations studied here lie at least 450 bp to the right of the startpoint for leftward *oop* transcription. Furthermore, deletion of the 123-bp Hpa II subfragment does not impair the replicator function of an *imm*λ-*ori*λ fragment cloned in Charon 3 (*30*). The deleted subfragment includes the *oop* promoter and all of the *oop* coding region except 14 bp corresponding to the 3′ end of the transcript (*8, 12*). Thus, *oop* RNA appears to be dispensable for λ DNA replication.

There remains a *cis* requirement for rightward transcription in the replication region, whether the replicator is at its usual location or on a transposed fragment (*2, 20*). Normally this RNA synthesis initiates at p_R, but mutants with new rightward promoters, such as *c*17 or the *ri*^{c} mutants, are able to bypass the requirement for transcription from p_R (*20, 30a, 31*). Remarkably, two of the *ri*^{c} mutations have been mapped to the right of the Eco RI site in gene *O* (*2, 18*) and thus do not promote rightward transcription of the region affected by the *ori*$^-$ mutations. This remains an enigma requiring further study.

KATHERINE DENNISTON-THOMPSON
DAVID D. MOORE, KELLEY E. KRUGER
MARK E. FURTH
FREDERICK R. BLATTNER
Laboratory of Genetics and *McArdle Laboratory, University of Wisconsin, Madison 53706*

References and Notes

1. D. D. Moore, K. Denniston-Thompson, M. E. Furth, B. G. Williams, F. R. Blattner, *Science*, **198**, 1041 (1977).
2. M. E. Furth, F. R. Blattner, C. McLeester, W. F. Dove, *ibid.*, p. 1046.
3. The structures of KK1 and KK2 are shown in Fig. 1. These phages were constructed to facilitate purification of small *imm-ori*λ Eco RI fragments, containing the λ replication origin, by removing extraneous Eco RI sites. KK1 was constructed by crossing λ*b*538*s*RI3°c1am509*nin5QSR*80 (from B. Williams) with λ*h*80*att*80*imm*434*QSR*80 [M. Wu, S. Ghosh, M. Willard, J. Davison, H. Echols, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 589] to obtain the structure shown in Fig. 1. The KK2 was constructed by B. Williams by crossing λKH100*P*am80 [F. Blattner, M. Fiandt, K. Haas, P. Twose, W. Szybalski, *Virology* **62**, 458 (1974)] with λ*b*189*bio*1*nin5QSR*80. Each of these phages has only two closely spaced Eco RI sites, defining the ends of the *imm*434-*ori*λ or *imm*λ-*ori*λ fragment, and can be used as a source of the fragment for purification by differential centrifugation (*1*).
4. Ava I (K. Murray and A. Morrison, unpublished observations) was a gift of G. G. Peters. Bgl II (G. Wilson and F. Young, unpublished observations) was a gift of T. Kelly. Hae II [R. Roberts, J. Breitmeyer, N. Tabachnek, P. Myers, *J. Mol. Biol.* **91**, 121 (1975)], Hinc II (J. Olson, P. Myers, R. Roberts, unpublished observations), Hha I (R. Roberts, P. Myers, A. Morrison, K. Murray, *J. Mol. Biol*, in press), Alu I (K. Subramanian and S. Weissman, unpublished observations), and Mbo I [R. Gelinas, P. Myers, R. Roberts, *J. Mol. Biol.* **114**, 169 (1977)]. DNA to be digested with Mbo I was propagated in *Escherichia coli* GM33, which has a defect in deoxyadenosine methylation (*dam*$^-$), to allow the enzyme to digest the DNA completely [M. Marinus and N. Norris, *J. Mol. Biol.* **85**, 209 (1976)]. Taq I [S. Sato, C. Hutchison, J. Harris, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 542 (1977)] was a gift of S. Sato and was used at 65°C. Hph I [D. Kleid, Z. Humayun, A. Jeffrey, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 293 (1976)] was a gift of R. Roberts, P. Myers, and M. Zabeau. Hpa II [P. Sharp, W. Sugden, J. Sambrook, *Biochemistry* **12**, 3055 (1973)] Eco RI [R. Yoshimori, thesis, University of California (1971); J. Hedgpeth, H. Goodman, H. Boyer, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3448 (1972)] was purified by the method described by B. Williams, F. Blattner, S. Jaskunas, M. Nomura [*J. Biol. Chem.* **252**, 7355 (1977)]. Digestions with all enzymes except Eco RI (*1*) were carried out at 37°C in 6 m*M* tris-HCl, *p*H 7.9, 6 m*M* $MgCl_2$, 6m*M* dithiothreitol. All enzymes were obtained from New England Bio-Labs, Beverly, Mass., unless otherwise specified.
5. Purified 2.75 percent *imm*434-*ori*λ fragments were end-labeled at Eco RI ends by exchange with the use of [γ-^{32}P]ATP (adenosine triphosphate) and T4 polynucleotide kinase (New England Bio-Labs). The [γ-^{32}P]ATP was prepared according to (*6*). End-labeling reactions were carried out at 37°C for 30 to 60 minutes in 50 m*M* tris-HCl, *p*H 7.6, 10 m*M* dithiothreitol, 5 m*M* $MgCl_2$, 0.05 m*M* ADP (adenosine diphosphate), with 0.5 mc of [γ-^{32}P]ATP (1 μ*M*) and three units of T4 polynucleotide kinase. Fragments la-

beled at the 5′ end were purified by Sephadex G-50 chromatography in a 0.3*M* sodium acetate, 0.01*M* tris-HCl, *p*H 7.9, 0.001*M* EDTA system, and concentrated by precipitation with ethanol.
6. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
7. Low molecular weight DNA fragments were resolved by electrophoresis at < 10 volt/cm in gels containing 10 percent polyacrylamide (acrylamide: *N,N'*-methylenebisacrylamide = 25 : 1) in one-half strength TBE [0.05*M* tris-borate, *p*H 8.3, 0.001*M* EDTA, A. Peacock and C. Dingman, *Biochemistry* **7**, 668 (1968)] with 25 percent glycerol. Glycerol improves band sharpness dramatically.
8. J. Dahlberg and F. Blattner, in *Virus Research*, C. Fox and W. Robinson, Eds. (Academic Press, New York, 1973), p. 533.
9. M. Rosenberg, personal communication.
10. D. Kleid, Z. Humayun, A. Jeffrey, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 293 (1976).
11. M. Rosenberg, B. de Crombrugghe, R. Musso, *ibid.*, p. 717.
12. G. Scherer, G. Hobom, H. Kössel, *Nature (London)* **256**, 117 (1977).
13. V. Pirotta, personal communication.
14. E. Schwarz, G. Scherer, G. Hobom, H. Kössel, personal communication.
14a. G. Scherer, G. Hobom, H. Kössel, personal communication.
15. B. Westmoreland, W. Szybalski, H. Ris, *Science* **163**, 1343 (1969).
16. D. Hewish, K. Smith, J. Spencer, *Virology* **74**, 363 (1976).
17. A. Rambach, *ibid.* **54**, 270 (1973).
18. W. Dove, H. Inokuchi, W. Stevens, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 747.
19. W. Szybalski *et al.*, *J. Cell. Physiol.* **74** (Suppl. 1), 33 (1969).
20. H. Inokuchi, W. Dove, D. Freifelder, *J. Mol. Biol.* **74**, 721 (1973).
21. N. Kleckner, *Virology* **79**, 174 (1977).
22. M. Schnös and R. Inman, *J. Mol. Biol.* **55**, 31 (1971); A. Kozinski and B. Lindqvist, personal communication.
23. The λ *ori* region, defined by the *ori*$^-$ mutations, has two dPy · dPu stretches (18 bp and 18 of 21 bp) in which the pyrimidines are on the *r* strand. The origin of *Col* E1 replication, defined by the site at which deoxyribonucleotides are added onto an RNA primer in vitro, is 20 bp away from a dPy · dPu stretch (21 of 24 bp) with purines on the transcribed strand [J.-I. Tomizawa, H. Ohmori, R. Bird, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1865 (1976)]. The origin region for phage fd DNA replication, defined by an in virto RNA primer, contains a dPy · dPu stretch (29 of 32 bp) with pyrimidines on the transcribed strand (C. Gray, R. Sommer, C. Polke, E. Beck, H. Schaller, personal communication). The pyrimidine tract CTC_3 lies in the region of the *A* gene dependent gap in φX174 at the start of RF to RF replication [S. Eisenberg, B. Harbers, C. Hours, D. Denhardt, *J. Mol. Biol.* **99**, 197 (1975)].
24. F. Sanger, G. Air, B. Barrell, N. Brown, A. Coulson, J. Fiddes, C. Hutchison, P. Slocombe, M. Smith, *Nature (London)* **265**, 687 (1977); P. Baas, H. Jansz, R. Sinsheimer, *J. Mol. Biol.* **102**, 633 (1976); S. Eisenberg, B. Harbers, C. Hours, D. Denhardt, *ibid.* **99**, 197 (1975).
25. R. Fisher, *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
26. We believe that the deletions observed in the nucleotide sequences for *r*93, *r*99, and *r*96 correspond to the *ori*$^-$ lesions. The *ori*$^-$ mutants were isolated without mutagenesis and occur with frequencies inconsistent with multiple mutations. The deletions could not have arisen during cloning, because they are present in the *ori*$^-$ phages λ*N*am7*N*am53c1857*r*93 and λ*N*am7*N*am53c-1857*r*99. The O protein made from the DNA of these phages in vitro is smaller than that made from DNA of the *ori*$^+$ parent phage when measured by SDS polyacrylamide gel electrophoresis [J. Yates and M. Furth, unpublished results, using techniques described in J. Yates, W. Gette, M. Furth, M. Nomura, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 689 (1977)]. The size of each deletion is consistent with the changes in molecular weights of the O protein made in vitro and of DNA restriction fragments. Finally, the recombinational map of the *ori*$^-$ mutations is consistent with the order of the sequenced deletions (*r*93-*r*99-*r*96), indicating that the deletions are responsible for the *ori*$^-$ pheonotype of these mutants (*2*).
27. J. Szpirer, *Mol. Gen. Genet.* **114**, 297 (1972); M. Furth, C. McLeester, W. Dove, in preparation; L. Lambert, R. Thomas, R. Monnat, W. Szybalski, personal communication.
28. By the criterion of heteroduplex pairing, the region of DNA encoding the COOH-terminal half of the O protein appears to be largely conserved in all five lamboid coliphages studied [M. Simon, R. Davis, N. Davidson, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 313; M. Fiandt, Z. Hradecna, H. Lozeron, W. Szybalski, *ibid.*, p. 329]. Complementation analysis shows that *P* gene function can be provided to λ by each of the other four lamboid phages [W. Dove, *Annu. Rev. Genet.* **2**, 305 (1969); J. Szpirer and P. Brachet, *Mol. Gen. Genet.* **108**, 78 (1970)]. In φ80/λ hybrid phages such as λ*imm*80*hy*42 the *O* gene (*27*) and O protein iteself have been shown to be hybrid (J. Yates and M. Furth, unpublished observations). There is genetic evidence for direct interaction between O and P proteins [J.-I. Tomizawa, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 549; M. Furth, C. McLeester, W. Dove, in preparation; (*2*)] and between P protein and the *E. coli* replication apparatus [C. Georgopoulos and I. Herskowitz, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 553; H. Saito and H. Uchida, *J. Mol. Biol.* **113**, 1 (1977)].
29. S. Hayes and W. Szybalski, in *DNA Synthesis and Its Regulation*, M. Goulian, P. Hanawalt, C. Fox, Eds. (Benjamin, Menlo Park, Calif., 1975), p. 486.
30. B. Williams, M. Furth, K. Kruger, D. Moore, K. Denniston-Thompson, F. Blattner, in preparation.
30a. W. Dove, E. Hargrove, M. Ohashi, F. Haugii, A. Guha, *Jpn. J. Genet.* **44** (Suppl. 1), 11 (1969).
31. H. Nijkamp, W. Szybalski, M. Ohashi, W. Dove, *Mol. Gen. Genet.* **114**, 80 (1971).
32. This is paper 2172 from the Laboratory of Genetics of the University of Wisconsin and paper No. 7 in the series "Charon phages for DNA Cloning." Paper No. 6 is (*2*). Supported by NIH grant GM21812 to F.R.B., NIH grant CA-07175 to McArdle Laboratory (W. F. Dove), an NSF predoctoral fellowship and an NIH training grant to McArdle Laboratory (M.E.F.), NIH training grant T32 CA09075 (K.D.-T.), and NIH training grant 144-J825 (D.D.M.). We thank W. F. Dove for support and advice, Nigel Godsen for critical reading of the manuscript, Ed Kopetsky and Brenda Dierschke for technical assistance, R. Roberts for hospitality in his laboratory, A. Maxam for providing the detailed sequencing procedures and A. Honigman for discussions concerning AvaX. This work was done under NIH guidelines, which call for EKl, Pl.

19 July 1977; revised 20 September 1977

Expression in *Escherichia coli* of a Chemically Synthesized Gene for the Hormone Somatostatin

Abstract. *A gene for somatostatin, a mammalian peptide (14 amino acid residues) hormone, was synthesized by chemical methods. This gene was fused to the* Escherichia coli *β-galactosidase gene on the plasmid pBR322. Transformation of* E. coli *with the chimeric plasmid DNA led to the synthesis of a polypeptide including the sequence of amino acids corresponding to somatostatin. In vitro, active somatostatin was specifically cleaved from the large chimeric protein by treatment with cyanogen bromide. This represents the first synthesis of a functional polypeptide product from a gene of chemically synthesized origin.*

The chemical synthesis of DNA and recombinant DNA methods provide the technology for the design and synthesis of genes that can be fused to plasmid elements for expression in *Escherichia coli* or other bacteria. As a model system we have designed and synthesized a gene for the small polypeptide hormone, somatostatin (Figs. 1 and 2). The major considerations in the choice of this hormone were its small size and known amino acid sequence (*1*), sensitive radioimmune and biological assays (*2*), and its intrinsic biological interest (*3*). Somatostatin is a tetradecapeptide; it was originally discovered in ovine hypothalamic extracts but subsequently was also found in significant quantities in other species and other tissues (*3*). Somatostatin inhibits the secretion of a number of hormones, including growth hormone, insulin, and glucagon. The effect of somatostatin on the secretion of these hormones has attracted attention to its potential therapeutic value in acromegaly, acute pancreatitis, and insulin-dependent diabetes.

The overall construction of the somatostatin gene and plasmid was designed to result in the in vivo synthesis of a precursor form of somatostatin (see Fig. 1). The precursor protein would not be expected to have biological activity, but could be converted to a functional form by cyanogen bromide cleavage (*4*) after cellular extraction. The synthetic somatostatin gene was fused to the lac operon because the controlling sites of this operon are well characterized.

Given the amino acid sequence of somatostatin, one can design from the genetic code a short DNA fragment containing the information for its 14 amino acids (Fig. 2). The degeneracy of the code allows for a large number of possible sequences that could code for the same 14 amino acids. Therefore, the choice of codons was somewhat arbitrary except for the following restrictions. First, amino acid codons known to be favored in *E. coli* for expression of the MS2 genome were used where appropriate (*5*). Second, since the complete sequence would be constructed from a number of overlapping fragments, the fragments were designed to eliminate undesirable inter- and intramolecular pairing. And third, G·C-rich (guanine-cytosine) followed by A·T-rich (adenine-thymine) sequences were avoided since they might terminate transcription (*6*).

Eight oligonucleotides, varying in length from 11 to 16 nucleotides, labeled