Proc. Nat. Acad. Sci. USA
Vol. 69, No. 11, pp. 3479–3484, November 1972

Fine-Structure Deletion Map of the *Escherichia coli* L-Arabinose Operon

(lambda phage/mating/regulation/suppression)

ROBERT SCHLEIF

Graduate Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02154

Communicated by Charles Yanofsky, September 18, 1972

ABSTRACT **A fine-structure deletion map of the L-arabinose operon of *E. coli* was constructed by mapping deletion endpoints against point mutations. Of 350 independent deletions with average endpoint separation of ten nucleotides, 51 ended in the control region between the *C* and *B* genes, and the rest ended in the structural genes *A*, *B*, *C*, and *D*. If deletion endpoints are randomly distributed, the *C* and *B* genes are separated by about 510 nucleotides. Deletion endpoints and locations of point mutations in fact do appear randomly interspersed in the *C* and *B* genes, but no point mutations were found in the control region between them. Deletions were isolated with the aid of a heat-inducible λ phage inserted into leucine genes adjacent to the arabinose genes. A high-capacity mating technique was developed for rapidly generating fine structure maps from many deletions and point mutations.**

Properties of mutant strains frequently provide information on the regulation of operons and the action of their proteins. For example, deletions show that the lactose operon operator lies between the *lac* structural genes and a site essential to their transcription (1). Fine-structure genetic mapping and physical techniques showed that a mutant *lac* repressor lacking the first 42 amino acids can form a tetrameric structure that binds inducer normally, but does not bind to DNA (2).

The arabinose operon of *Escherichia coli* is a particularly interesting object for a detailed genetic analysis. Its structural genes *A*, *B*, and *D* (gene order *thr araDABC leu*) are controlled positively by arabinose and by the arabinose *C* protein *in vivo* (3), as well as *in vitro* (4), but the mechanism for this control is not known. Furthermore, the *B*, *A*, and *D* enzymes are well characterized and easy to purify (5–7). Finally, strains with deletions ending between *C* and *B* appear worthy of further study, since one such deletion strain possessed unusual regulatory properties (8).

As the distance between point mutations decreases, recombination frequencies become poor indicators of mutation order, due to multiple crossovers in localized regions (9, 10). Therefore, fine structure mapping may better be done by deletion mapping, where the experimental question is whether or not a functional gene can be reconstructed (11).

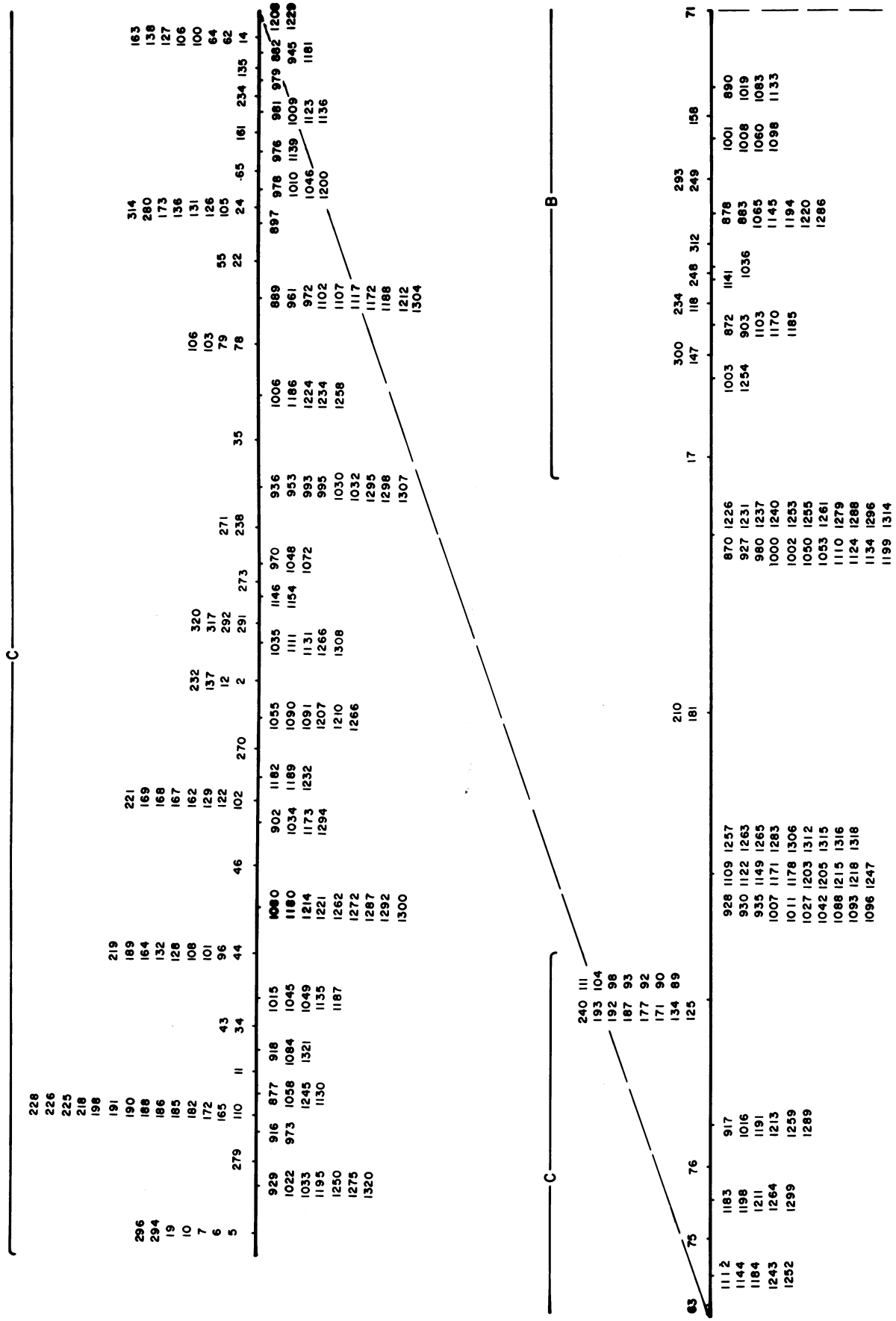
In this work, I used a strain in which a heat-inducible λ phage was inserted into the leucine region adjacent to the arabinose genes (12). Cells with deletions extending into the arabinose operon were greatly enriched among survivors after induction of the phage. These arabinose-deletion female strains were mapped by mating with strains containing episomes carrying arabinose point mutations.

METHODS

Deletion Isolation. Isolated colonies of RFS830 were grown to 3×10^9/ml in 100 ml of M9 medium (13), 1 μg/ml of thiamine, 50 μg/ml of L-leucine, and 0.2% L-arabinose (M9-B1-Leu-Ara) at 35°. This was added to 600 ml of YT broth (13) at 25° and shaken 24 hr at 42°; 0.05 ml of these cells was added to 1 ml of M9-B1-Leu medium containing 0.2% glycerol, 0.4% arabinose, and 2.5% ribitol, and grown 24 hr to stationary phase. Cells possessing functional *C* and *B* genes phosphorylate ribitol and do not grow (15). Ara$^-$ cells grow well. 200 of these cells, 95% Ara$^-$, were spread on tetrazolium arabinose plates and, after 20 hr at 35°, were replicated onto minimal B1-Leu-Ara plates spread with 2×10^7 cells of RFS867, a strain carrying F′*araD*$^-$. After 48 hr, colonies indicating Ara$^-$ on tetrazolium, but producing Ara$^+$ colonies on the minimal plate by recombination with the F′*D*$^-$, were picked. Only one deletion per original culture was kept.

Point Mutation Isolation and Initial Characterization. RFS740 was grown in M9-B1-Leu-Ara medium, and 10^8 cells were spread on solidified ribitol medium as above. After 2 days, the about 100 ribitol-resistant colonies were picked and purified on tetrazolium arabinose plates. Strains F′*ara-1* through F′*ara-24*, F′*ara-34* through F′*ara-61*, F′*ara-62* through F′*ara-65*, F′*ara-73* through F′*ara-79*, F′*ara-80* through F′*ara-141*, and F′*ara-142* through F′*ara-238* were isolated from cultures grown from different colonies. These strains were classified by complementation as Ara *A*$^-$, *B*$^-$, or *C*$^-$ by mating with strains RFS855, RFS859, and RFS1157. F′239 through 332 were isolated from RFS1165 containing an *ara D*$^-$ nonsense mutation that produced arabinose sensitivity due to accumulation of substrate for *D* enzyme. Isolated colonies of RFS1165 were grown to stationary phase in M9-B1-glycerol medium, and 0.1 ml was spread on minimal plates containing 0.2% glycerol and 0.4% arabinose. After 36 hr, the plates containing about 500 colonies were replicated onto two minimal arabinose plates, one spread with 2×10^9 $\phi 80 psu_{III}$, a nonsense-suppressing phage (16). After 20 hr of growth, the about 20 colonies growing only on the $\phi 80 psu_{III}$ plates were purified on tetrazolium arabinose plates. The phage segregates rapidly from cells and Ara$^-$ colonies are found. Only mutants of these nonsense strains mapping in different locations were kept from one culture.

Mapping. Strains grown overnight in 1 ml of M9–4% strength YT medium were diluted with 3 ml of YT medium–0.2% glucose, with pH adjusted to 7.9 to maximize mating. This was added with a sterile Labindustries Repipet. After 2 hr of growth, drops of males and females were added to rows and columns of a polypropylene block (21.6 × 31.2 × 1.9 cm, 0.95-cm diameter, 1.3-cm deep wells on 1.3-cm centers). Mating proceeded for 3–5 hr. Male–male mating was performed as above, except that the recipients were grown in YT medium for 30–36 hr before mating. Samples from 96


C
B
C

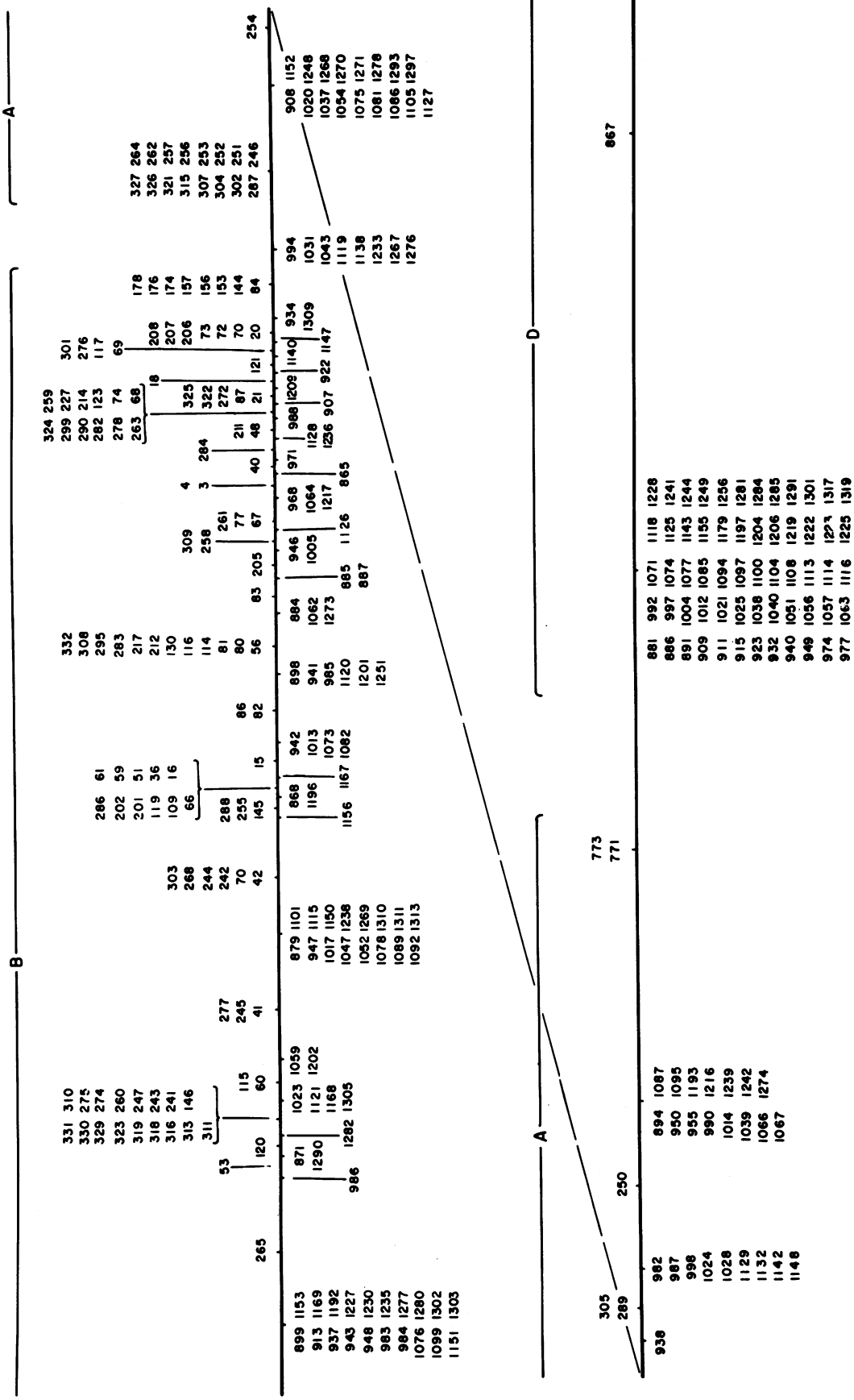


FIG. 1. Fine-structure map of the arabinose operon. The endpoints of deletions entering from the leucine region to the left as drawn on the map are shown below the line and the locations of point mutations are shown above the line. The map is in four segments, the first two on the *left page* and the second two on the *right page*, as connected by *dashed lines*. The numbers refer to strain numbers; those below the line are RFS X and above the line are RFS F'X.

TABLE 1. *Strains*

KS73	Hfr: Δ(*gal attλ bio uvrB*) *leu-73*	From Robert Weisberg (12). HfrH λCl_{857} inserted into *leucine*.
RFS740	F′: *thr*$^+$ *ara*$^+$ *leu*$^+$/*ara-498 thi lac-74*	Pasteur strain RV with *lac* deletion X74, Thi$^-$, *arabinose–leucine* deletion *498* (13) P1 transduced in, contains Thr$^+$ Ara$^+$ Leu$^+$ episome KLF1 (14), No. F-101 in Genetic Stock Center, Dept. of Microbiology, Yale University, New Haven, Conn. 06510.
RFS817	F$^-$: *ara-498 lac-74 thi str*	Thi$^-$, Str^r.
RFS830	F$^-$: *leu-73* Δ(*gal attλ bio uvrB*) *lac-74 str lam-830*	Ara$^+$ Str^r product of KS73 x-RFS817, made resistant to λ.
RFS855	F$^-$: *araA855 thr leu lac-74, recA111, nalA111 tsx*	Pasteur strain RV with *lac* deletion X74, Ara$^-$ by nitrosoguanidine, Thr$^-$ and Leu$^-$ from C600 by P1; Rec$^-$ and nalidixic acid-resistant from Hfr111 of Joel Kirshbaum, T6^r.
RFS859	F$^-$: *araC859* otherwise like RFS855	
RFS867	F′: *araD54 thr*$^+$ *leu*$^+$/*araD54 leu* tsx-867	Episome KLF 1, AraD$^-$ (13) Leu$^-$ from Strain C600 by P1, T6-resistant.
RFS1157	F$^-$: *araBF15 leu lac-74 recA111 nalA111 tsx*	Like RFS855, but Ara$^-$ from F′*ara-15* by P1.
RSF1165	F′*araD591 thr*$^+$ *leu*$^+$/*ara-498 lac-74 thi nalA111*	Episome KLF1, arabinose-sensitive but suppressible by su_{III} to Ara$^+$, Thi$^-$, nalidixic acid-resistant, from Hfr111 of Joel Kirshbaum.

wells at a time were spotted onto 23 × 33-cm Pyrex trays containing 300 ml of solidified arabinose-B1 medium–1% strength YT medium with an aluminum plate through which machine screws (1.9 cm, 10/32, ground plane) protruded to enter the wells. The spotting plate was sterilized by flushing with hot tap water, dipping in 95% ethanol 30 sec, and blowing dry. After 24–48 hr of growth, the trays were rapidly scored with a typewriter, row by row, by typing + for growth of a spot and − for no growth. Using these methods, one person could routinely handle and score four trays of matings per day—1456 different crosses.

RESULTS

Source of deletions

Powerful selection techniques must be used to isolate the very infrequent cells containing a deletion entering the arabinose operon (frequency $<10^{-10}$ per cell). After induction of a strain containing a heat-inducible λ phage inserted into the leucine region adjacent to the arabinose genes, about 5×10^{-3} cells are Ara$^-$ and 2×10^{-5} possess deletions ending within the arabinose operon. The desired *ara* deletions were found by selecting for Ara$^-$ cells and testing for a part of the end gene, *D*, in each strain. By these means, one cell from 10^{11} could be isolated with a deletion entering the arabinose genes from the leucine region and ending within the arabinose operon.

Deletions could be generated by incorrect λ excisions. However, Table 2 shows that individual cultures in which the phage was heat induced contained widely differing numbers of Ara$^-$ survivors, indicating that the deletions are generated *before phage induction*. If the deletions were generated at the time of induction, all the cultures would be equivalent and the numbers of deletions would then be distributed in accord with the Poisson distribution, with variance equalling the mean. A great majority of cells possessing a prophage at the time of induction must be killed even though deletions may be generated that dismember the phage and extend into the arabinose operon.

Source of point mutations

Large numbers of point mutations were isolated. All point mutations were isolated on an episome to facilitate mating with the female deletion strains. Sensitivity of Ara$^+$ cells to ribitol was used in the selection of the C^- and B^- mutations numbered F′*ara-1* through F′*ara-238*. For F′*ara-239* through F′*ara-332*, an F′D^-_{amber} mutation, conferring arabinose sensitivity, allowed selection of A^-, B^-, and C^- mutations. The D^- mutation is suppressible by su^+_{III}, and only mutants resistant to arabinose and suppressible to Ara$^+$ by su^+_{III} were chosen. The latter strains thus possess two nonsense mutations in the arabinose operon, one in genes *A*, *B*, or *C* and one in the *D* gene.

Mapping

Sets of male point mutants were crossed with sets of female deletion strains. Each map assignment was checked for consistency with all other crosses in the set, and each strain was mapped at least twice. Consistent map locations were obtained in all cases; the resultant map is shown in Fig. 1. This drawing was made by allotting each deletion a fixed length of map. All deletions in a subgroup not separated by point mutations are presented by a single hashmark. Endpoints of the deletions entering from the left are shown below the line and point mutations are shown above. Fewer point mutations were mapped in the *A* gene than in the *B* and *C* genes; thus, the deletions ending in *A* are not well subdivided. Although each deletion is an independent clone, some point mutations may be siblings (see *Methods*).

Distribution of deletion and point mutations

The large number of deletions and point mutations mapped allows a comparison to be made between their distributions within the *C* and *B* genes, and the distribution expected if both deletions and point mutation locations were randomly distributed. The frequency of n deletions ending between

two adjacent point mutations, R(n,M,N), for M different deletions and N different point mutations will be considered.

The probability that two adjacent point mutations are separated by a distance x in a unit interval is $P(x,N) \approx \lim_{\Delta x \to 0} N[(1 - \Delta x)^N]^{x/\Delta x} = N\exp(-Nx)$. The probability that n deletions end in an interval x is $Q(x,n,M) = M!/[n!(M - n)!]x^n(1 - x)^{M-n}$, so that $R(n,M,N) = \int_0^1 PQdx$. Letting $P(x,N) = N(1 - x)^N$, which is a close approximation at the maximum of the integrand when N is large, one finds $R(n,M,N) = N/(N + M)[M/(N + M)]^n$.

The number M of independent deletions ending within *C* and *B* can be counted; however, the number N of different independent point mutations cannot be counted, since point-mutation "hot spots" (11) and siblings most likely generated multiple copies of some mutants. The number N, of independent locations of point mutations, nonetheless, can be calculated from the number of point mutations separated by at least one deletion, N', for $N' = \sum_{n=1}^{n=\infty} NR(n,M,N) = N[1 - R(0,M,N)]$. Substitution from above for R(n,M,N) yields $N = N'M/(M - N')$. In our case, M = 221, N' = 61, and thus N = 84. Elimination of the n = 0 term, which corresponds to siblings and point mutations at the same point, as well as to point mutations at different points but not separated by deletions, and normalization so that $\sum_{n=1}^{n=\infty} R(n,M,N) = 1$, yields $R(n,M,N) = N/M[M/(N + M)]^n$, with N given as above. Fig. 2 shows the theoretical and experimental distributions taken from the map of Fig. 1. They agree well for $n < 15$. However, near the right end of *B* is a set of 20 deletions not separated by any point mutations. Fig. 2 shows that this is improbable and, therefore, this region may be a deletion "hot spot" or a region in which few mutations are phenotypically expressed because many different aminoacid substitutions are tolerated in the corresponding part of the protein. Except for this region, deletion endpoints and the different locations of point mutations seem to be randomly distributed in the *C* and *B* genes.

TABLE 2. *Luria-Delbrück (17) fluctuation test on Ara⁻ survivors of phage induction*

Ara⁻ colonies	No. of plates
0	39
1	26
2	11
3	6
4	2
5	3
7	3
9	1
11	1
20	1
28	1

Average per plate = 1.78; variance = 14.9. Average per culture = 71.2; bacteria per culture = 1.2×10^{10}. Mutation rate, 5.8×10^{-10} per cell doubling. Isolated colonies of RFS830 were grown to 3×10^9/ml in 4 ml of YT medium, and 0.1 ml was spread on tetrazolium arabinose plates that had been warmed to 42°; the inoculated plates were incubated at 42° for 20 hr. The number of Ara⁺ survivors also fluctuated widely.

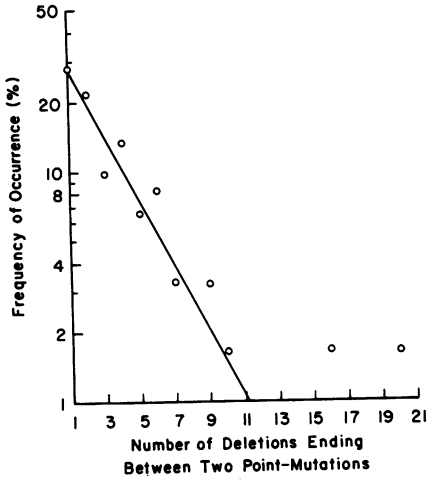


FIG. 2. The frequency of occurrence of *n* deletions ending between two point-mutations. The *solid line* is the frequency predicted as in the text and the *points* are the experimental data determined from the map in the *C* and *B* genes.

DISCUSSION

Notable in this deletion map is the large fraction of deletions ending in the control region between the *ara C* and *B* genes. If deletion endpoints are randomly distributed, then, based the masses of the *C* and *B* gene monomers of 30,000* and 50,000 daltons, respectively, and the fraction of deletions ending in or between the two, this area is about 510 nucleotides long.† However, deletions could preferentially end in the control region. A special structure, possibly associated with regulation of the operon, could account for such a "hot spot" effect. An estimate of about 510 nucleotides in the *ara* control region may be compared with data for the *lac* and *trp* operons of *E. coli*, the *his* operon of *Salmonella typhimurium*, and phage λ indicating about 100, 2000, 200, and 200 nucleotides in the respective control regions (1, 20–22).

No point mutations, but many deletions, were found between *C* and *B*. Two mutations on episomes F'*ara-181* and F'*ara-210* were mapped in this region, but they were deletions of the *A* and *B* genes. The absence of point mutations in the control region implies that changes of a single base do not render the cells sufficiently arabinose negative to be isolated by my selective methods.

No deletions have been found ending in a nontranslated region at the beginning of the *C* gene, for to the right of the

* Unpublished experiments with Jack Greenblatt show the *C* protein activity assayed in a cell-free protein synthesizing system sediments at 4 S and, thus, that the active form of *C* is about 60,000 daltons. For these calculations *C* is taken to be a dimer, since *in vivo* experiments show a cooperative arabinose induction curve, predicting at least a dimeric structure for *C* (18, 19).

† The probability that *n* out of *N* random deletions end within a subinterval of length *x* within a unit-length interval is $N!/[n!(N - n)!]x^n(1 - x)^N$. This function is sharply peaked for *x* corresponding to 510 nucleotides, and the probability that the length of the region is 260 relative to the probability that it is 510 is 1.6×10^{-4}.

rightmost deletions are two nonsense mutations, F'*ara-294* and F'*ara-296*, showing that they are translated and, therefore, that the rightmost deletions are within the translated portion of the *C* gene. Physiological studies indicate that the *C* gene is transcribed from right to left (23) and, therefore, if the *C* gene truly is oriented in this direction, a nontranslated RNA polymerase binding site for transcription of *C* may be small.

I thank Cathy Becker for help with the crosses. This work was supported in part by the National Institute of General Medical Sciences, NIH (Research Grant GM-18277 and Career Development Award K4-GM-38797 to R.F.S.). Publication No. 866 from the Graduate Department of Biochemistry, Brandeis University.

1. Miller, J., Ippen, K., Scaife, J. & Beckwith, J. (1968) *J. Mol. Biol.* **38**, 413–420.
2. Platt, T., Weber, K., Ganem, D. & Miller, J. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 897–901.
3. Sheppard, D. & Englesberg, E. (1967) *J. Mol. Biol.* **25**, 443–454.
4. Greenblatt, J. & Schleif, R. (1971) *Nature New Biol.* **233**, 166–170.
5. Lee, N. & Bendet, I. (1967) *J. Biol. Chem.* **242**, 2043–2050.
6. Patrick, J. & Lee, N. (1968) *J. Biol. Chem.* **243**, 4312–4318.
7. Lee, N., Patrick, J. & Masson, M. (1968) *J. Biol. Chem.* **243**, 4700–4705.
8. Englesberg, E., Squires, C. & Meronk, F. (1969) *Proc. Nat. Acad. Sci. USA* **62**, 1100–1107.
9. Amati, P. & Meselson, M. (1965) *Genetics* **51**, 369–379.
10. Sadler, J. & Smith, T. (1971) *J. Mol. Biol.* **62**, 139–169.
11. Benzer, S. (1961) *Proc. Nat. Acad. Sci. USA* **47**, 403–415.
12. Shimada, K. & Weisberg, R. (1972) *J. Mol. Biol.* **63**, 483–503.
13. Schleif, R. (1969) *J. Mol. Biol.* **46**, 185–196.
14. Low, B. (1968) *Proc. Nat. Acad. Sci. USA* **60**, 160–167.
15. Katz, L. (1970) *J. Bacteriol.* **102**, 593–595.
16. Russell, R., Abelson, J., Landy, A., Gefter, M. Brenner, S. & Smith, J. (1970) *J. Mol. Biol.* **47**, 1–13.
17. Luria, S. & Delbrück, M. (1943) *Genetics* **28**, 491–511.
18. Doyle, M., Brown, C., Hogg, R. & Helling, R. (1972) *J. Bacteriol.* **110**, 56–65.
19. Schleif, R. (1969) *J. Mol. Biol.* **46**, 197–199.
20. Hiraga, S. (1969) *J. Mol. Biol.* **39**, 159–179.
21. Fankhauser, D. (1969) *Genetics* **61**, s17.
22. Blattner, F., Dahlberg, J., Boettiger, J., Fiandt, M. & Szybalski, W. (1972) *Nature New Biol.* **237**, 232–236.
23. Schleif, R., Greenblatt, J. & Davis, R. (1971) *J. Mol. Biol.* **59**, 127–150.