

- I. ELECTRON MICROSCOPE HETERODUPLEX ANALYSIS
OF DNA SEQUENCES IN F-PRIME FACTORS
- II. ELECTRON MICROSCOPE STUDIES OF λ AND
MU PROPHAGES
- III. AN ELECTRON MICROSCOPE STUDY OF SINDBIS
VIRUS RNA

Thesis by

Ming-Ta Hsu

In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1974

(submitted October 22, 1973)

ii

To my mother

ACKNOWLEDGMENTS

I wish to express my deep gratitude to Professor Norman Davidson not only for his patient guidance in research but particularly for his always warm concern for my well-being.

I also wish to express my gratefulness to Drs. P. A. Sharp and E. Ohtsubo for setting enthusiastic examples in scientific research; Mr. H. J. Kung for collaboration in research; Prof. and Mrs. J. H. Strauss for their advice and generosity in providing laboratory facilities for the preparation of Sindbis virus; Professors J. Vinograd and G. Attardi for their always helpful advice ; Mr. P. F. Koen for his instruction and assistance in using the electron microscope; and to all those wonderful people in ND's group for their help and advice in either scientific or personal matters.

ABSTRACT

This thesis is composed of three parts. Part I is concerned with studies of the DNA sequences of several F-prime factors and the sequence relations among them using electron microscope heteroduplex methods. It was found that the DNA sequence of the F factor can be roughly divided into three regions: 1) a region about one-fourth of the molecule which is concerned with the fertility functions of the F factor. 2) a region about one-third of the molecule which is rich in A+T sequence and contains the sequences used to interact with the bacterial chromosome for the integration or excision of F factor DNA. 3) a region which contains the genes for autonomous replication and female phage resistance and the structural element for conjugal transfer.

The structure of the bacterial DNA carried in several classical episomes, F100, F152, F8 and some of their derivatives was studied extensively. Bacterial markers between fep and uvrB were analyzed both genetically and physically. A method was developed to reconstruct the original episomes from their deletion variants. The results confirm the history that F100 and F152 were derived from the same Hfr. The formation of a new episome, F80, from F8 suggested that there is a hot spot in the E. coli chromosome for the recombination with F sequence between 93.2 and 94.5/0.0F.

In part II, the structures of λ and Mu prophages and Mu phage DNA were studied. The λ prophage carried in an F-prime factor was found by electron microscope heteroduplex analysis to be circularly permuted relative to the vegetative viral DNA. On the other hand, Mu prophage DNA was shown to be collinear with the viral DNA. The integrated Mu prophage DNA was used as a marker for physical mapping of bacterial genes in E. coli.

Sequence heterogeneity in Mu phage and prophage DNA's was also studied. The G loop heterogeneity was found to be present in both phage and prophage DNA's and was shown to be due to sequence inversion. The heterogeneous split ends sequences were found to be absent in the several prophages studied.

Part III contains an electron microscope study of viral RNA of Sindbis virus and a method for mapping poly A sequences in RNA molecules. Under weak denaturing conditions Sindbis virus RNA appears in circular form with a double stranded "handle" of about 250 nucleotides long. This implies that Sindbis RNA contains complementary sequences at or near the ends of the molecule.

A technique using glyoxal as a denaturing agent for mapping polyA sequences in RNA was developed. Glyoxal attaches to guanine base irreversibly and thus removes

the secondary structure of RNA without inhibiting the renaturation capacity of polyA sequences in the molecule. A polyA sequence was found at or near one end of Sindbis RNA by this method.

TABLE OF CONTENTS

<u>Part</u>	<u>Title</u>	<u>Page</u>
I	ELECTRON MICROSCOPE HETERO- DUPLEX ANALYSIS OF DNA SEQUENCES IN F-PRIME FACTORS.	1
Chapter 1	Electron Microscope Heteroduplex Studies of Sequence Relations among Plasmids of <u>E. coli</u> . I. Structure of F-prime Factors	2
Chapter 2	Electron Microscope Heteroduplex Studies of Sequence Relations among Plasmids of <u>E. coli</u> . III. Structures of F100, F152, F8 and Their Derivatives.	36
Chapter 3	Electron Microscope Analysis of Partial Denaturation of F Factor DNA.	106
II.	ELECTRON MICROSCOPE STUDIES OF λ AND MU PROPHAGES	130
Chapter 1	On the Structure of Prophage λ	131
Chapter 2	Structure of Inserted Bacterio- phage Mu-1 DNA and Physical Mapp- ing of Bacterial Genes by Mu-1 DNA Insertion.	137

<u>Part</u>	<u>Title</u>	<u>Page</u>
II Chapter 3	Electron Microscope Hetero- duplex Study of the Hetero- geneity of Mu Phage and Pro- phage DNA.	143
III	AN ELECTRON MICROSCOPE STUDY OF SINDBIS VIRUS RNA.	170
PROPOSITIONS		
1	A New Method for Isolation of λ Transducing Phages Carrying Any Part of <u>E. coli</u> Chromosome.	196
2	Application of the Diaminoben- zidine Technique for the <u>in</u> <u>situ</u> Localization of Integrated Viral Genome and Single-Copy Genes.	204
3	Applications of an Inverse Immu- noadsorption Technique in Isola- ting and Analyzing Proteins	211
4	A Method for Eliminating Fluores- cence Background in Laser Raman Spectroscopy	216
5	A New Method for Genetic Mapping in Mammalian Cells	221

Part I

ELECTRON MICROSCOPE HETERODUPLEX ANALYSIS
OF DNA SEQUENCES IN F-PRIME FACTORS

Chapter 1

Electron Microscope Heteroduplex
Studies of Sequence Relations among
Plasmids of E. coli. I. Structure
of F-prime Factors.

Electron Microscope Heteroduplex Studies of Sequence Relations among Plasmids of *Escherichia coli*

I. Structure of F-prime Factors

PHILLIP A. SHARP†, MING-TA HSU, EIICHI OHTSUBO
AND NORMAN DAVIDSON

*Department of Chemistry, California Institute of Technology
Pasadena, Calif. 91109, U.S.A.*

(Received 18 February 1972, and in revised form 10 June 1972)

The sequence relations between some bacterial F-prime factors in *Escherichia coli* have been determined by observing, in the electron microscope, the pattern of duplex and single-stranded regions in heteroduplexes consisting of one strand from one episome and the complementary strand from another.

All of the F-primes studied have a short piece of F missing. They appear to have been formed by excision from their respective Hfr's by a type I process (Scaife, 1967), leaving a piece of F remaining in the chromosome. The piece of F missing is the same in *Flac* and in the close relatives F450 and F_{1a}, although the former F-prime was derived from Hfr2 and the latter two from Hfr3. Thus, there is evidence for hot spots on F for insertion at different places in the *Escherichia coli* K12 chromosome to give Hfr's and hot spots for excision to give F-primes. However, a different sequence of F is missing in the episome F₈.

The physical structure of a number of F₈ episomes has been studied. In particular, we have studied the physical and genetic structures of a set of deleted F₈ episomes, prepared by P1kc transduction of the original F₈, and mostly deleted in some of the transfer genes. Thus, the various genetic markers have been physically mapped. A general model for the structure of F is proposed.

1. Introduction

An F-prime factor is a circular DNA molecule harbored within a bacterium. It contains some bacterial DNA and some or all of the fertility factor DNA molecule, F. As is well known, F-primes are formed in two steps from a male bacterium harboring the F factor: (1) insertion of F into the bacterial chromosome to give an Hfr; (2) a subsequent aberrant excision of a circular DNA molecule containing the essential genes of F and some bacterial genes (Scaife, 1967; Campbell, 1969; Jacob & Wollman, 1961).

These processes are shown in Figure 1. The point on F which is the origin for transfer of DNA into a recipient female is depicted as an arrow between sequences *U* and *T*. If an F DNA molecule is transferred by mating of an F⁺ male bacterium with a recipient female, the sequences of DNA transferred are in the order *U, V, W, . . . F, K, T*. The sequences of DNA transferred by mating of the Hfr are in the order *U, V, W, α, β, . . . ψ, ω, X, Y, Z, . . . T*, where Greek letters represent bacterial genes; *α* and *β* are markers transferred early in a mating, *ψ* and *ω* are transferred late,

† Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 11724, U.S.A.

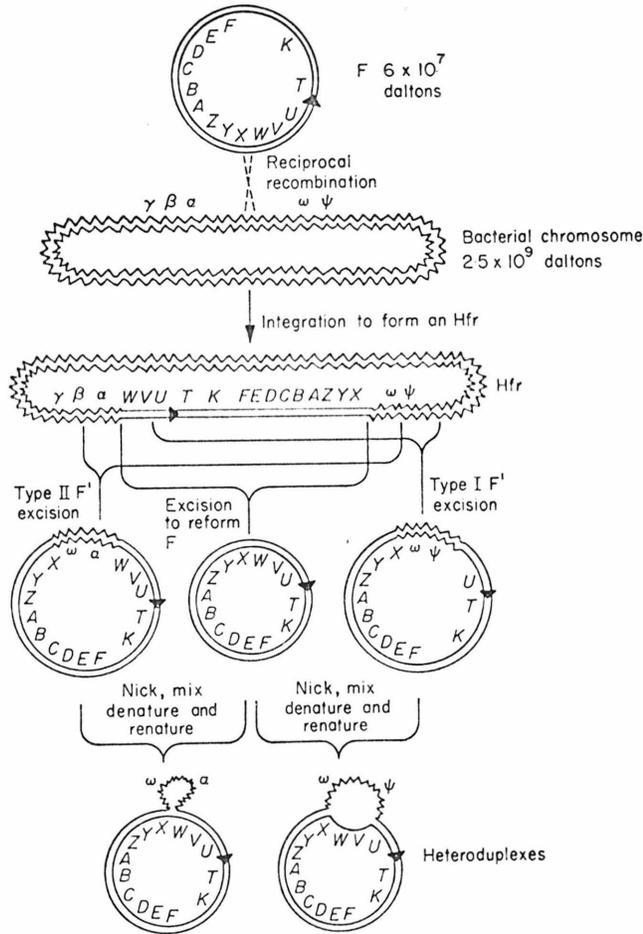


FIG. 1. Processes leading to the formation of an Hfr and subsequent excision to form F-primes. The structures expected in heteroduplexes are also shown. For further discussion, see text.

followed by the remaining F sequences $X, \dots T$. Since the property of maleness is itself transferred late in an Hfr mating, some of the essential genes of F must be included within $X, \dots T$. The symbols $A, \dots Z$ in Figure 1 are arbitrary, except that A to F represent mapped essential transfer cistrons (Ohtsubo, 1970). The origin of transfer and the site of recombination indicated in the Figure are also somewhat arbitrary.

By the reverse excision process, the Hfr can revert to the original F factor and the original bacterial chromosome. By an aberrant excision, presumably an illegitimate recombination (Franklin, 1971), a circular F-prime, carrying F sequences and some bacterial sequences, is formed. If, as in type II excision (Fig. 1), the illegitimate recombination occurs between bacterial sequences on both sides of F, the resulting F' will contain all of F and both early and late bacterial sequences depicted as a and ω . If the illegitimate recombination occurs between an F sequence and a bacterial sequence, as illustrated as type I excision in Figure 1, the F' will be missing a part of F

shown as *W*, *V* and carry only late bacterial markers. (The notation I and II in Fig. 1 are those introduced by Scaife (1967, Fig. 5, p. 618).) To be fertile, this F' must carry the known transfer cistrons *A* to *F* and other unidentified essential groups. (Another type I process not depicted in Fig. 1 would be an illegitimate recombination between, for example, a point between *X* and *Y* on F and a point between β and γ on the bacterial chromosome, thus giving an F' deleted in the F sequence *X* and carrying the early bacterial markers α and β .)

The genes in F code for the various characteristics of fertile male bacteria: ability to transfer DNA; superinfection immunity; autonomous replication; presence of pili; sensitivity to various classes of male-specific phage; etc. The transfer functions have to some extent been genetically analyzed by complementation (Ohtsubo, Nishimura & Hirota, 1970; Ohtsubo, 1970; Achtman, Willetts & Clark, 1971; Willetts & Achtman, 1972; Ihler, Achtman & Willetts, 1972; T. Miki, personal communication).

In the present work we have applied the electron microscope heteroduplex method (Davis & Davidson, 1968; Westmoreland, Szybalski & Ris, 1969; Davis, Simon & Davidson, 1971) to the study of the structure and sequence relationships of various F' factors. The problems to which we have addressed ourselves are: (1) determination of the molecular length of an F' and its content of F sequences, and of bacterial sequences; (2) determination of whether a particular F' has the structure resulting from a type I or type II excision process (Fig. 1); (3) mapping of the recombination site on F (between *X* and *W* in Fig. 1) at which F inserted to form the Hfr from which the F' was derived. By comparing different F'-primes, we can thus ask whether the insertion of F occurs at only one site on F or whether there are a number of sites on F at which it is capable of recombining with the bacterial chromosome. (4) For a type I process, physical mapping of the aberrant excision site on F (between *U* and *V* in Fig. 1); (5) the physical mapping of bacterial genes present in a particular F' when possible (for example, by renaturation with a specialized transducing phage carrying known bacterial genes). (6) Physical mapping of the transfer cistrons of F and of bacterial markers by a study of the deleted F₈ primes constructed by Ohtsubo (1970). (7) Observation of other, previously unsuspected, structural features of a given F or F'.

The general principles of the electron microscope heteroduplex method and its particular application to the solution of problems (1) and (2) above are illustrated in Figure 1. Covalently closed, circular F and F' molecules are nicked so that there are very few doubly nicked strands. A mixture of the nicked F and F' molecules are subjected to strand dissociation by NaOH, reneutralized and allowed to renature. Some of the renatured molecules consist of one strand of F and one strand of the F'. These are the heteroduplexes of interest. The molecules are mounted for electron microscopy by the formamide modification of the basic protein-spreading technique (Westmoreland *et al.*, 1969; Davis *et al.*, 1971) so that both single- and double-stranded regions are seen as extended filaments. Thus, topological features of the molecule can be recognized. Note that because of topological constraints, two complementary circular strands cannot renature fully with each other. A linear strand and a circular strand can renature. Complementary linear strands resulting from random nicking of circular molecules can renature and will most probably give a circular structure when the two breaks are not close to each other.

If the F' is formed by a type II process, the region of non-homology will be an insertion loop as indicated in Figure 1 with the single-stranded loop consisting of the inserted bacterial DNA and the circular duplex part being the complete F molecule.

If the F' is formed by a type I process, the heteroduplex with F will show a substitution loop (Fig. 1). One of the single-stranded branches is the sequence *V, W* of F which is missing in the F'; the other single-stranded branch contains the bacterial DNA of the F'.

The extensions or modifications of these methods for the study of problems (3) to (7) will be described as the results are presented.

2. Materials and Methods

(a) Bacterial strains and bacteriophage

Bacterial strains used in this study are described in Table 1. Bacteria were grown without cloning unless otherwise specified. Bacteriophage λ c26 and λ b2b5c are laboratory

TABLE 1
Bacterial strains

Strain	Plasmid	Plasmid genotype ^a	Chromosomal genotype	Source and reference
B583	F450	<i>uvrB</i> ⁺ <i>att</i> ^λ <i>bio</i> ⁺ <i>gal</i> ⁺ <i>tra</i> ⁺	Δ (<i>gal-att</i> ^λ - <i>chlA</i>)	D. Freifelder ^{d, o}
W1485	F	<i>tra</i> ⁺	<i>thi</i>	D. Freifelder ^{d, f}
YS40F ⁺	F	<i>tra</i> ⁺ (F transferred from W1485)	<i>his pro str</i>	D. Holinski ^g
W6	F	<i>tra</i> ⁺	<i>met</i>	E. Lederberg ^f
JC1553(<i>Flac</i>)	<i>Flac</i>	<i>lac</i> ⁺ <i>tra</i> ⁺	<i>leu</i> ₁ <i>his</i> ₁ <i>arg</i> ₆ <i>lac</i> _{1,4} <i>mal</i> ₁ <i>str recA</i>	D. Holinski ^g
MR44	F _{1a}	<i>bio</i> ⁺ <i>att</i> ^λ <i>gal</i> ⁺ <i>supE</i> <i>tra</i> ⁺	<i>galE galT recA spc</i>	E. Signer ^{e, f}
W3101(F' <i>gal</i>)	F Δ (0-14.5) ^b	<i>tra</i> ⁺	<i>gal</i> ₁	A. D. Kaiser
ND3	F Δ (0-14.5)	<i>tra</i> ⁺	<i>gal</i> ₆ <i>trp arg recA str</i>	This paper
PA106(F' ₂ <i>gal</i>)	F Δ (0-14.5)	<i>tra</i> ⁺	<i>thr leu thi lys lac gal xyl recA</i>	J. Gross, ex. F. Jacob, via P. Broda
W4520	F ₈ (1) ^o	<i>gal</i> ⁺ <i>aroG</i> ⁺ <i>nadA</i> ⁺ <i>tra</i> ⁺	<i>met</i>	Y. Hirota & P. H. A. Sneath ^{f, 1}
JE5303	F ₈ (1) ^o	<i>gal</i> ⁺ <i>aroG</i> ⁺ <i>nadA</i> ⁺ <i>tra</i> ⁺	<i>gal</i> ₆ <i>trp arg recA str</i>	E. Ohtsubo ¹
JE3100	F ₈ (2) ^o	<i>gal</i> ⁺ <i>aroG</i> ⁺ <i>nadA</i> ⁺ <i>tra</i> ⁺	<i>thr leu gal</i> ₂ <i>lac pil fla str</i>	E. Ohtsubo, Y. Nishimura & Y. Hirota ^k
JE3513	F ₈ (N33)	<i>gal</i> ⁺ <i>aroG</i> ⁺ <i>nadA</i> ⁺ <i>traA</i> ⁻	<i>thr leu gal</i> ₂ <i>lac pil fla str</i>	E. Ohtsubo, Y. Nishimura & Y. Hirota ^k
JE3445	F ₈ (N73)	<i>gal</i> ⁺ <i>aroG</i> ⁺ <i>nadA</i> ⁺ <i>traB</i> ⁻	<i>thr leu gal</i> ₂ <i>lac pil fla str</i>	E. Ohtsubo, Y. Nishimura & Y. Hirota ^k
N23-53			F ⁻ <i>gal</i> ₆ <i>trp arg recA str</i>	H. Ogawa, K. Shimada & J. Tomizawa ¹
PL225 <i>recA</i>			F ⁻ Δ (<i>nadA-aroG-gal</i>) <i>recA str</i>	K. Mizuuchi and T. Fukasawa ^m and this paper
For other F ₈ <i>gal</i> derivatives, see Figs 11 and 12				

^a The symbol *tra*⁺ means possessing all the usual properties of male bacteria: transfer, presence of pili, sensitivity to male specific phage, etc. (Ohtsubo *et al.*, 1970).

^b F Δ (0-14.5) is a deleted F, missing the sequences from 0 to 14.5 kb; see text.

^c F₈(1) and F₈(2) were both called F₈ in Ohtsubo *et al.* (1970) and Ohtsubo (1970); see text.

^d Freifelder, Folkmanis & Kirschner (1971).

^e The history of F450 and F_{1a} as we understand them are as follows. An episome F₁ *gal att*^λ *bio* was derived from Hfr 3 by Jacob & Wollman; Hfr 3 is an Hfr transferring *purE* as an early marker and *gal*, *att*^λ and *bio* as late markers (p. 162 of Jacob & Wollman, 1961; see also Liedke-Kulke

stocks derived from the late Dr Jean Weigle. λdg LBT2 was a kind gift from Dr Allan Campbell; it is described as $\lambda dg1^-$ by Kayajanian & Campbell (1966). M13 phage was a gift from Dr Arleen Forsheit.

(b) *Media*

Bacteria were usually grown in a Tryptone broth containing 1 μ g thiamine/ml. Davis minimal agar plates were used for genetic experiments, with either 0.2% glucose or 1% galactose as a carbon source (Davis & Mingioli, 1950). Eosin-methylene blue indicator agar plates were used for *gal* and *lac* markers.

(c) *Isolation of plasmid DNA's*

Bacterial cultures were grown to late logarithmic phase of approximately 10^9 cells/ml. in 2 l. of medium (see (b) above). The cells were centrifuged and washed twice in Tris/EDTA/saline buffer (0.05 M-NaCl, 0.05 M-Tris, 0.005 M-EDTA, pH 8.5) at 0°C. The lysis procedure is a slight modification of that of Bazarol & Helinski (1968). The bacterial pellet was resuspended in 100 ml. of a spheroplast-forming mixture, containing 1 mg lysozyme/ml., 100 μ g RNase/ml. and 0.1 g sucrose/ml. in Tris/EDTA/saline buffer, and incubated at 37°C for 10 min. This suspension was chilled in an ice-bath; lysis of the cells was completed by adding 50 ml. of 2% sodium lauryl sarcosine solution in Tris/EDTA/saline buffer. The resulting viscous solution was mixed (and the DNA sheared) by forcing 50-ml. batches through a 50-ml. disposable syringe (orifice int. diam. 1.5 mm) in 20 to 25 sec. This is repeated 4 to 6 times. The sheared lysate was brought to room temperature and titrated to pH 12.3 by adding 4 to 6 ml. of 4 N-NaOH, and stirring vigorously; pH was measured with a Beckman Research model pH meter and a Sargent S30070-10 pH 0 to 14 miniature combination electrode, after standardizing with pH 10 buffer. The solution was then reneutralized to pH 8.5 by the addition of 2 M-Tris-HCl. The rationale of this procedure is that by titration to pH 12.3 and reneutralization, linear and open circular DNA's are denatured, but a closed circular DNA remains native. Excess denatured DNA was removed by a bulk nitrocellulose procedure modified from that of Cohen & Miller (1969). Bulk nitrocellulose, 100 g, (Hercules, $\frac{1}{4}$ sec cubed) that had been ground by pestle and mortar and washed twice with Tris/EDTA/saline buffer was rotated with the DNA solution for 1 hr at 4°C. The nitrocellulose was removed by centrifugation and the DNA in the remaining 180 ml. supernatant was pelleted onto a 3-ml. saturated CsCl shelf for 16 hr at 15 krev./min in an SW25.2 rotor.

& Kaiser, 1967). E. Wollman sent a bacterial stock bearing $F_1 gal att^A bio$ to Professor A. D. Kaiser. The episome was transferred successively into several different hosts by investigators in Kaiser's laboratory, and finally into SA224, thus giving rise to strain B583. SA224 contains a deletion spanning *galE* to *chlA*; not all of *galK* is deleted; *aroA* is not deleted (S. Adhya, personal communication). F_{1s} was derived by Signer, Beckwith & Brenner (1965) from an $F_1 gal att^A bio$ -bearing strain sent to them by E. Wollman.

[†] Our understanding of the pedigree of F and of the several F-primes studied here is as follows. Strain 58-161, $F^+ bio met$, was derived from strain 58 (or W4), $F^+ bio$, which was derived from *E. coli* K12, F^+ , in both cases by X-ray treatment in the laboratories of Tatum & Lederberg. W1485 $F^+ lac$ was directly derived from *E. coli* K12, F^+ , by u.v. treatment by Tatum & Lederberg. 58-161 was sent to Jacob & Wollman who selected, as spontaneous mutants, the Hfr strains, Hfr₂ from which *Flac* was derived and Hfr₃ from which $F_1 gal$ and $F_2 gal$ were derived (p. 162 of Jacob & Wollman, 1961; Liedke-Kulke & Kaiser, 1967). 58-161 spontaneously reverted to W6, $F^+ met$ which was treated with u.v. to give W3208, Hfr₈, *met* by Lederberg. There is some question as to whether this strain is a true Hfr. W4520, $F_8^+ met$ was derived from W3208 by Hirota & Sneath (1961) by several matings. The above history is in part based on personal communications from E. Lederberg and Y. Hirota.

^{*} Bazarol & Helinski (1968).

^h Clark & Margulies (1965).

[†] Hirota & Sneath (1961).

[‡] Ohtsubo (1970).

^k Ohtsubo, Nishimura & Hirota (1970).

^l Ogawa, Shimada & Tomizawa (1968).

^m Mizuuchi & Fukasawa (1969).

The lower 8 ml. from each SW25.2 tube was pooled and ethidium bromide was added to give a concentration of 500 $\mu\text{g/ml}$. The density was adjusted to 1.55 with CsCl and the DNA banded by centrifugation for 24 hr at 40 krev./min in an SW50.1 rotor. The lower bands (closed circular molecules) were collected, pooled and rebanded for 20 hr in an SW50.1 rotor. The resulting lower band for a typical F preparation contained 10 to 30 μg of DNA in 0.5 ml. and is the DNA used in all further experiments. Ethidium bromide was removed by dialysis against 10 g of Dowex W-X (100 to 200 mesh) (in 50 ml. of 0.8 M-NaCl, 0.05 M-Tris, 0.01 M-EDTA, pH 8.5) for 24 hr at 4°C. The DNA solution was further dialyzed against three 250-ml. changes of 0.25 M-NaCl, 0.05 M-Tris, 0.01 M-EDTA (pH 8.5) and stored at 4°C for daily use. For long-term storage, DNA was left in the ethidium bromide/CsCl solution as collected from the gradient.

For most of the episomes isolated as closed circular molecules, that is, as lower bands in ethidium bromide/CsCl, we observed, by electron microscopy, a low frequency, approx. 1%, of interlocked circles (catenanes). This was true in either *recA*⁺ or *recA*⁻ hosts. After X-ray nicking, some of these molecules appeared as an intermediate band.

(d) *X-ray nicking of DNA*

The closed, circular DNA, as collected from the second banding above without dilution, was adjusted to 0.002 M-histidine and irradiated with 220 kV X-rays to give approximately 50% nicking of the closed circles (Freifelder, 1968; Freifelder & Meselson, 1970). We find that in the ethidium bromide/CsCl medium a dose of 780 R will convert approximately 0.5 of the F DNA from closed to open circles. After nicking, the DNA is rebanded in ethidium bromide/CsCl, and the upper band collected for heteroduplex formation.

Table 2 presents an analysis of the relation between the average number of nicks per single strand (\bar{n}), the fraction, $1-p_{00}$, of molecules converted to an open circular form II, and the fraction of strands with different numbers of nicks. The calculation assumes that nicking is a random process.

In general, one wishes to maximize the number of intact strands that move into the open circular band, $p_0(1-p_0)+p_1$ (see the legend to Table 2). At the same time one wishes

TABLE 2
Statistics of random nicking

\bar{n}	p_0	p_1	p_{00}	$1-p_{00}$	$1-p_0-p_1$	$F_{0, \text{II}}$	$F_{1, \text{II}}$	$F_{2, \text{II}}$	$p_0(1-p_0)+p_1$
0.2	0.82	0.16	0.67	0.33	0.02	0.45	0.50	0.05	0.31
0.3	0.74	0.22	0.55	0.45	0.04	0.43	0.49	0.08	0.41
0.4	0.67	0.27	0.45	0.55	0.06	0.40	0.49	0.11	0.47
0.6	0.55	0.33	0.30	0.70	0.12	0.32	0.47	0.17	0.58
0.8	0.45	0.36	0.20	0.80	0.19	0.31	0.45	0.24	0.61
1.0	0.37	0.37	0.14	0.86	0.26	0.25	0.44	0.31	0.60

\bar{n} = Average number of nicks per strand; randomly distributed among strands by assumption.

p_1 = Probability that a strand receives i nicks; $p_1 = \bar{n}^i e^{-\bar{n}}/i!$ (Poisson distribution).

p_{00} = Fraction of duplexes with no nicks in either strand; therefore, $p_{00} = p_0^2 = e^{-2\bar{n}}$; p_{00} is the fraction of duplexes in the closed circular band.

$1-p_{00}$ = Fraction of duplexes in the open circular band.

$1-p_0-p_1$ = Fraction of all strands with two or more nicks.

$F_{0, \text{II}} = p_0(1-p_0)/(1-p_0^2)$ = Fraction of strands in the open circular band that are un nicked.

$F_{1, \text{II}} = p_1/(1-p_0^2)$ = Fraction of strands in the open circular band with one nick.

$F_{2, \text{II}} = (1-p_0-p_1)/(1-p_0^2)$ = Fraction of strands in the open circular band with two or more nicks.

$p_0(1-p_0)+p_1$ = Fraction of all strands that have zero or one nick and that are in the open circular band.

to avoid having many strands with two or more nicks, since these do not give complete heteroduplexes. The calculations in the Table suggest that these conflicting requirements are reasonably well satisfied for values of \bar{n} between 0.3 and 0.8, corresponding to 45 to 80% conversion to open circular molecules.

(e) *Electron microscope mapping procedures*

Molecular lengths of nicked duplex plasmid DNA's were determined by length measurements of molecules spread by the aqueous basic protein film technique. The concentrated DNA solution isolated from the CsCl gradient and dialyzed as described above was diluted into 0.5 M-ammonium acetate, 100 μ g cytochrome *c*/ml. (pH 7.0) and spread on to 0.25 M-ammonium acetate; films were picked up on a parlodion-coated grid, stained and shadowed as previously described (Davis *et al.*, 1971). Circular λ c26 molecules or (less frequently) ϕ X174 RFII molecules were present as an internal length standard.

For heteroduplex formation approx. 0.1 μ g of DNA of each of the two plasmids was added to 70 μ l. of 0.3 M-NaOH in a small test tube. This causes denaturation and strand dissociation. After 3 to 5 min 30 μ l. of 1 M-Tris-HCl and 100 μ l. of 0.2 M-trisodium EDTA (pH 8.5) were added and the solution dialyzed at room temperature for 2 hr against 50 ml. of 70% formamide, 0.25 M-NaCl, 0.10 M-Tris, 0.01 M-EDTA (pH 8.5). This is the renaturation step. (Unless otherwise specified, a Tris buffer is made by mixing Tris-HCl and Tris-OH to give the desired pH and diluted to give the given total concentration of Tris/EDTA stock solutions are prepared from disodium EDTA plus NaOH to the desired pH, then added to the final solution. A solution specified to contain 0.1 M-Tris, 0.01 M-EDTA (pH 8.5) thus contains approximate concentrations: (Tris-OH) = 0.067 M, (Tris⁺) = (Cl⁻) = 0.033 M, (Na⁺) = 0.03 M, ((H)EDTA⁻³) = 0.01 M; the cation concentration, which is the important parameter for renaturation, is thus approx. 0.063 M.) The renaturation step was followed by 2 hr of dialysis at 4°C against 50% formamide, 0.1 M-Tris, 0.01 M-EDTA, pH 8.5. Aqueous cytochrome *c* solution was added to give a spreading solution, usually 45% formamide, 0.1 M-Tris, 0.01 M-EDTA, 50 μ g of cytochrome *c*/ml. and spread on to a hypophase of 17% formamide, 0.01 M-Tris, 0.001 M-EDTA, pH 8.5. In these heteroduplex experiments, single-stranded ϕ X174 was added as a single-strand length standard and renatured intact plasmid homoduplexes of either type served as duplex length standards.

Length measurements are reported in daltons and/or in units of 1000 bases or base-pairs (kilobases). The value of $30.8 (\pm 1.0) \times 10^6$ daltons has been recommended for the molecular weight of the sodium salt of λ DNA (Davidson & Szybalski, 1971); this corresponds to 46,500 (± 1500) base pairs (average residue weight for a base pair = 662). We find a ratio of 0.113 (± 0.003) between the length of ϕ X174 RFII DNA and λ DNA. Thus, the molecular weight of ϕ X-RF is calculated as $3.48 (\pm 0.1) \times 10^6$ daltons, corresponding to 5250 (± 170) base pairs. We find that the length of the F factor from W1485 and YS40 (F⁺) is 2.03 (± 0.05) times λ or 17.95 (± 0.4) ϕ X-RF, corresponding to 94,500 (± 2400) base pairs. If the molecular weight of the *E. coli* chromosome is taken between 2.5 and 3×10^9 daltons (Cairns, 1963) or 3.8 to 4.5×10^3 kb \dagger , 1 min on the genetic map of *E. coli* K12 is 42 to 50 kb.

In heteroduplex experiments a mixture of two related DNA's, AA' and BB' is denatured and renatured. The resulting renatured mixture contains some remaining single-stranded molecules, the original homoduplex molecules AA' and BB', and the heteroduplexes AB' and A'B. Under the formamide conditions of mounting both single- and double-stranded DNA's appear in the electron microscope as somewhat extended filaments, but the single-stranded filaments are usually perceptibly thinner and a little kinkier than are the double-stranded filaments.

Heteroduplex molecules contain duplex segments, insertion and/or deletion loops and substitution loops. The duplex segments are identified as regions of sequence homology and the several kinds of single-stranded segments as regions of non-homology. Some heteroduplexes are incomplete because of single-strand breaks; some heteroduplexes are complete but are not sufficiently well spread out to be suitable for measurements. Thus, the number of measurable complete molecules in an electron microscope grid is usually small.

\dagger Abbreviation used: kb, kilobases, thousands of base pairs.

These complete heteroduplexes, however, serve to define the over-all structure clearly. Length measurements on different single- or double-stranded segments of incomplete structures (partial structures), which occur with a much higher frequency, can then be used with confidence to accumulate sufficient data for good statistics on the length of a given feature.

The general principles of length measurements and of their standard deviations, using internal standards, are discussed by Davis *et al.* (1971). In the present study, we have usually made about 10 measurements of any given feature—a duplex or single-stranded segment or an entire molecule. A minimum of 5 measurements was accepted for those cases when the population of molecules with measurable features on the electron microscope grid was sparse. Illustrative details for a particular set of measurements are given in the legend to Figure 3.

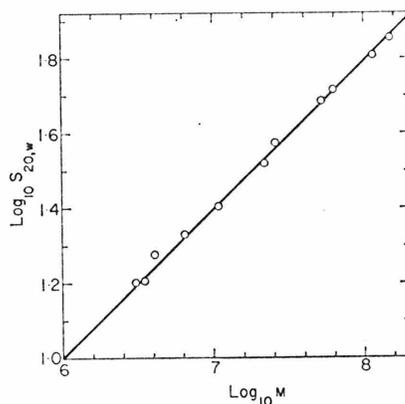


FIG. 2. Plot of the sedimentation coefficients of the open circular DNA form of the plasmids, as listed in Table 3. The least squares straight line through the data points is $\log_{10} S_{20,w} = 0.386 \log_{10} M - 1.29$. Alternatively, if an equation of the form $S_{20,w}^{-2.7} = b_1 M^{0.445}$ as suggested by the analysis of Crothers & Zimm (1965) and of Gray, Bloomfield & Hearst (1967) is fitted to the data, we find $S_{20,w}^{-2.7} = 0.01759 M^{0.445}$. The theoretical equation of Gray *et al.* (1967) is $S_{20,w}^{-2.7} = 0.01759 M^{0.445}$.

A single-stranded loop emanating from a point in a duplex and returning to the same point is due either to an insertion or a deletion of a sequence from one of the two DNA molecules. It is not always possible to tell which; we then refer to such a feature as an I-D loop.

We usually wish to know whether all the DNA molecules in a given preparation are identical by the heteroduplex criterion (the sample may appear to be homogeneous by the cruder criterion of duplex length distribution). The sample is denatured, renatured, and examined by electron microscopy for duplexes with non-homology regions; such a preparation has been self-renatured.

How much sequence homology is required for a segment to appear as duplex after spreading for electron microscopy under given conditions? This problem has been discussed by Davis & Hyman (1971). We estimate that the conditions of mounting from 45% formamide, described above, correspond to a denaturing condition about 25°C below T_m for a 50% G+C DNA. A 1% change in formamide concentration at constant ionic strength corresponds to approximately 0.7°C lowering in T_m (McConaughy, Laird & McCarthy, 1969). It has been further estimated that about 1.5% of randomly distributed mismatched bases result in a lowering of T_m of 1°C (Laird, McConaughy & McCarthy, 1969). If the renaturation conditions correspond to $T_m - 25^\circ\text{C}$, this would be the T_m for a duplex with the average base composition of the DNA's under consideration (50% G+C in the present case) with 37% of randomly distributed mismatched bases. Thus, in a segment which appears as duplex in a typical heteroduplex experiment, the mating strands could con-

coivably be only partially homologous with somewhat more than 63% (100—37) sequence complementarity. In the present instance, we are studying closely related DNA molecules derived from a common ancestor but which have had the opportunity to diverge by storage and cultivation in different laboratories, and by an occasional mutagenic treatment. It seems probable that the degree of sequence homology between segments which do re-nature to form a duplex is probably much higher than the 63% limit stated above.

3. Results and Discussion

(a) *Molecular weights and sedimentation coefficients*

The molecular lengths of the several plasmid DNA's studied are listed in Table 3. Sedimentation coefficients of some of the plasmids in the open circular form, as

TABLE 3
*Molecular weights and sedimentation coefficients of open circular
DNA molecules*

DNA	Mol. wt $\times 10^{-6}$ (daltons)	Molecular length kb	$S_{20,w}^0$ ^a
Polyoma	3.0 ^b	4.6	16.0 ^b
ϕ X174	3.4 ₆ ^c	5.2 ₅	16.2 ^c
M13	4.0 ^d	6.1	18.9 ^d
PM2	6.5 ₆ ^e	9.9 ₂	21.4 ^f
Mitochondrial monomers HeLa	10.7 ^e	16.1	25.6 ^g
Mitochondrial dimers HeLa	21.4 ^e	32.3	33.1 ^g
λ b2b5c	25.6 ₅ ^e	38.7	37.4 ^h
λ c26	30.8	46.5	
λ b2b5c dimer	51.3	77.4	48.5
F	62.6 \pm 1.6 ^o	94.5	52.0 \pm 0.4 ^f
F450	114.1 \pm 2.7 ^o	172.1	64.7 \pm 0.4
F450(λ)	147.4 \pm 4.2 ^o	222.6	72.0 \pm 0.5
<i>Flac</i>	94.9 \pm 2.1 ¹	143.3	
F _{1a}	150.1 \pm 2.8 ¹	226.7	
F Δ (0—14.5)	53.0 \pm 1.6 ¹	80.0	
F ₈ (1)	77.8 \pm 2.5 ^o	117.3	
F ₈ (2)	73.6 \pm 2.0 ^{o,1}	111.2	
F ₈ (N33)	74.9 \pm 1.8 ^o	113.1	
F ₈ (P54)	63.9 \pm 1.2 ^o	96.5	
F ₈ (P43)	53.8 \pm 0.9 ^o	81.3	
F ₈ (P322)	47.5 \pm 0.9 ^o	71.8	
F ₈ (P17)	49.7 \pm 1.7 ^o	75.1	
F ₈ (P6)	42.9 \pm 1.1 ^o	64.8	
F ₈ (P376)	43.5 \pm 0.9 ^o	65.7	
F ₈ (P405)	53.0 \pm 1.7 ¹	80.0	
F ₈ (P77)	76.0 \pm 3.5 ^o	114.8	

^a Sedimentation coefficients were corrected to $S_{20,w}^0$ by the method of Bruner & Vinograd (1965).

^b Vinograd, Lebowitz, Radloff, Watson & Laipis (1965).

^c Burton & Sinsheimer (1965).

^d Ray, Preuss & Hofschneider (1966).

^e Calculated from contour length measurements relative to lambda DNA taking the molecular weight of the latter as 30.8 (\pm 1.0) $\times 10^6$ daltons, or 46.5 (\pm 1.5) kb.

^f W. Upholt, personal communication.

^g Clayton & Vinograd (1967).

^h Wang (1970).

¹ Calculated from measurements of heteroduplexes with F.

measured by band velocity sedimentation in 3 M-CsCl, are also listed in the Table. Some data from the literature are included for comparison. A plot of $\log_{10} S_{20,w}^{\circ}$ versus $\log_{10} M$ is shown in Figure 2. As discussed in the legend to the Figure, there is an excellent fit to either of the two semi-empirical equations which have been used to correlate molecular weights and sedimentation coefficients for DNA. Thus, our new data show that these correlations extend to a higher molecular weight range for open circular DNA molecules than has been previously studied.

(b) $F_8(N33)$

This episome has several unusual structural features which can serve as recognizable reference points in heteroduplexes of $F_8(N33)$ with other plasmids. We have accordingly found it useful in mapping the sequence relations of many other plasmids to study their heteroduplexes with $F_8(N33)$ as well as with F.

The relevant properties of $F_8(N33)$ and of some other F_8 episomes are described here. A systematic presentation of all of the structural data on F_8 episomes is given in sections (i) and (j) below.

$F_8(N33)$ is transfer-defective in the *A* complementation group of F; it carries the *gal* bacterial markers (Ohtsubo *et al.*, 1970) and also the *nadA* and *aroG* sequences (see sections (i) and (j)).

An electron micrograph of an $F_8(N33)$ /F heteroduplex is shown in Plate I(a). (It should be noted that the F used in all of our work is the molecule extracted from W1485. Its relation and probable identity to the original F in *E. coli* K12 and to some other F factors is discussed in a later section.) The structure of the heteroduplex as deduced from measurements of a number of such micrographs is shown schematically in Figure 3(a). In this representation of the structure, unprimed numbers and solid lines refer to F sequences; primed numbers and a sawtooth line refer to bacterial sequences carried by the F'.

The molecular length of $F_8(N33)$ is $113(\pm 3.1)$ kb. There is a large substitution loop, with single-stranded branches *bs* and *fs* of length $20.7(\pm 1.2)$ and $7.3(\pm 0.6)$ kb, respectively. There are, in addition, two small I-D loops, labeled *c* and *d* in Figure 3(a), of length $1.2(\pm 0.1)$ and $1.4(\pm 0.1)$ kb, respectively. Evidence to support the belief that these are insertions is given below. F itself has a molecular length of $94.5(\pm 2.4)$ kb. We therefore propose that $F_8(N33)$ has a structure with a piece of F of length 7.3 kb missing, and that the single-stranded branch, *bs*, is the bacterial DNA. The predicted duplex length of $F_8(N33)$ from this model is $94.5(\pm 2.4) - 7.3(\pm 0.6) + 20.7(\pm 1.2) + 1.2(\pm 0.1) + 1.4(\pm 0.1)$ or $110.5(\pm 3.0)$ kb, in agreement with the observed length of $113(\pm 3.1)$ kb.

In Figure 3(a) and Plate I(a), as in most Figures and Plates of heteroduplex structures, the junctions between F DNA and bacterial DNA, where substitution loops begin and end, are labeled *a* and *b*. Single-stranded bacterial and F DNA are labeled *bs* and *fs*, respectively. We assign co-ordinates to bacterial DNA sequences in the F-prime as primed numbers. We usually assign co-ordinates to sequences of the bacterial DNA by measuring distances in kb from an origin (*O'*) at one junction of the bacterial DNA with F DNA. However, in the particular case of $F_8(N33)$, the counterclockwise junction *a* is given the bacterial co-ordinate 5.9' for consistency with other F_8 plasmids as explained later. If the necessary information is available, the direction of increasing co-ordinates for bacterial DNA sequences corresponds to the clockwise direction on the conventional *E. coli* map (Taylor & Trotter, 1967). This arbitrary numbering

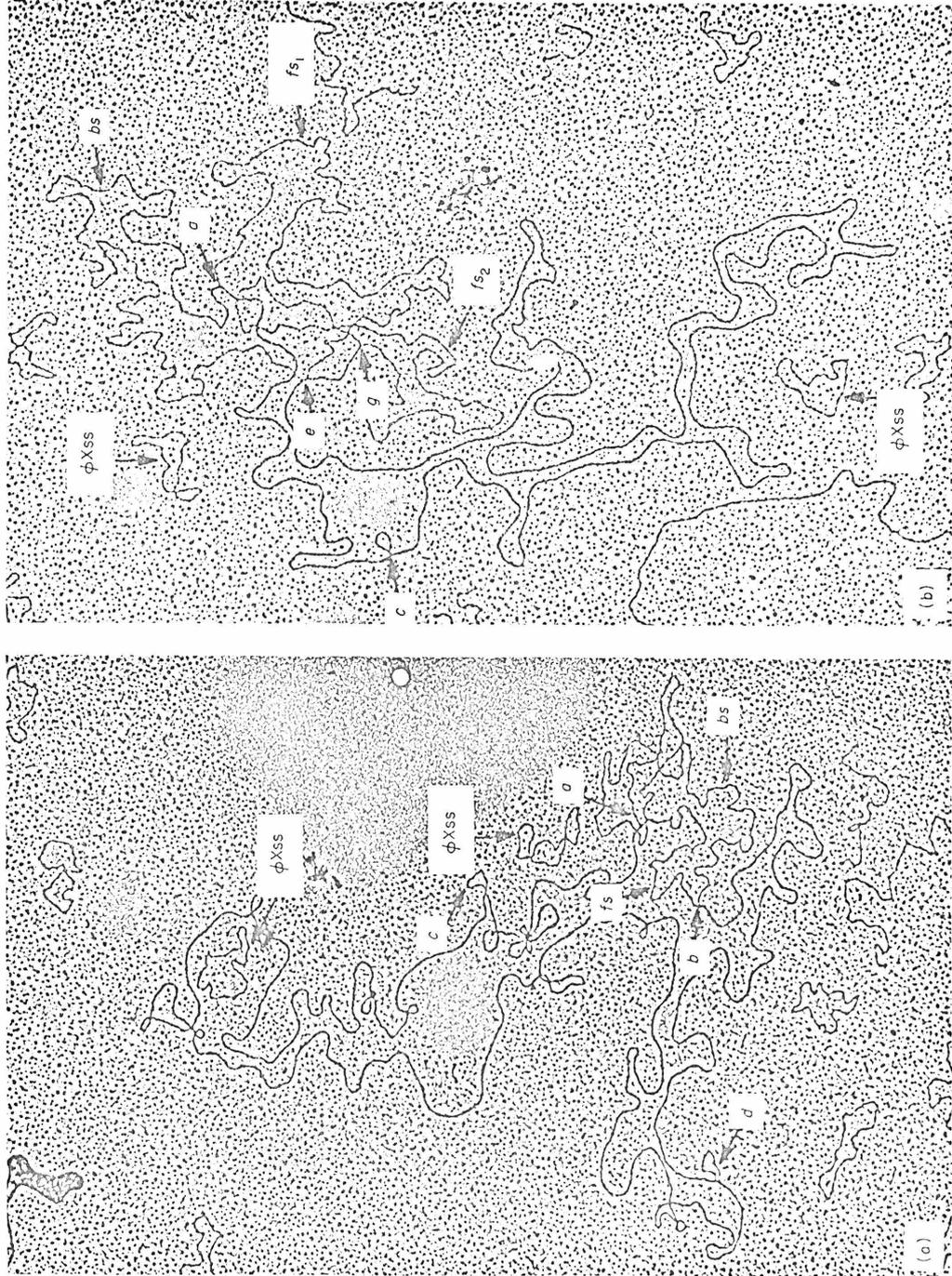


PLATE I. (a) $F_6(N33)/F$ heteroduplex. The schematic representation is shown in Fig. 3(a). Arrows *c* and *d* mark the two I-D loops at 91.0 and 34.9, respectively. Arrows *a* and *b* are the junctions of duplex DNA with single-stranded DNA at (5.9/8.5) and (26.6/15.8), respectively. Duplex F DNA, single-stranded F DNA, and single-stranded bacterial DNA are identified as *fd*, *fs*, and *bs*, respectively. Several single strands of ϕ N DNA are also marked. *ss*, Single-stranded. (b) Alternative structure of the $F_6(N33)/F$ heteroduplex in which the inverted repeat manifests itself. The important features of the structure are labeled in the micrograph and in the schematic representation in Fig. 3(b). These letter symbols are explained in the text.

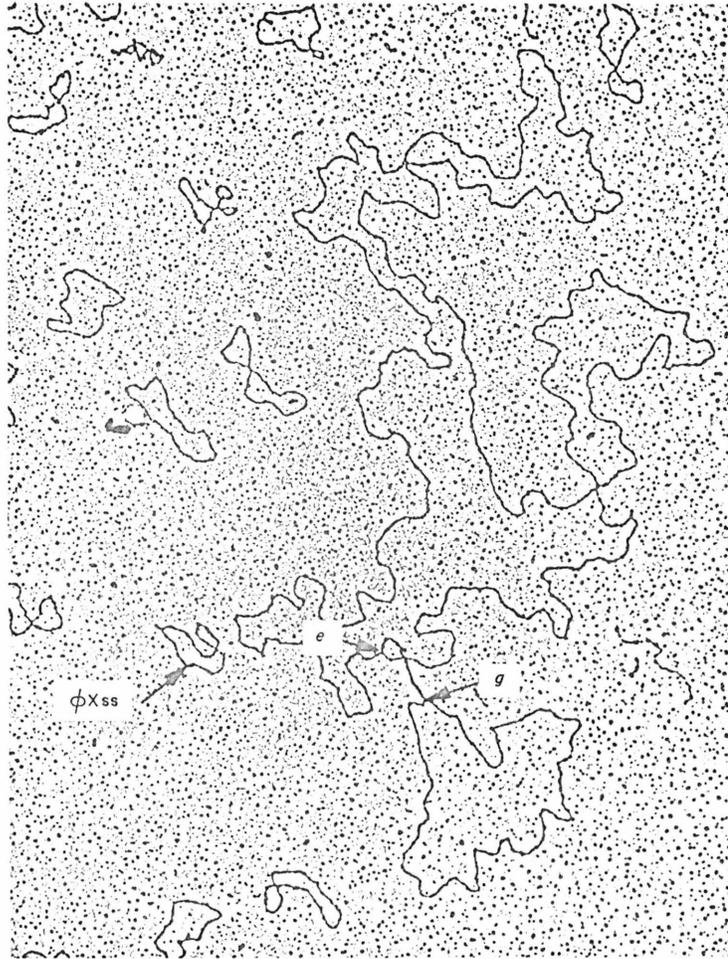


PLATE II. Electron micrograph of a single strand of $F_6(N33)$ DNA showing the duplex segment between points e and g due to the inverted repeat. ss, Single stranded.

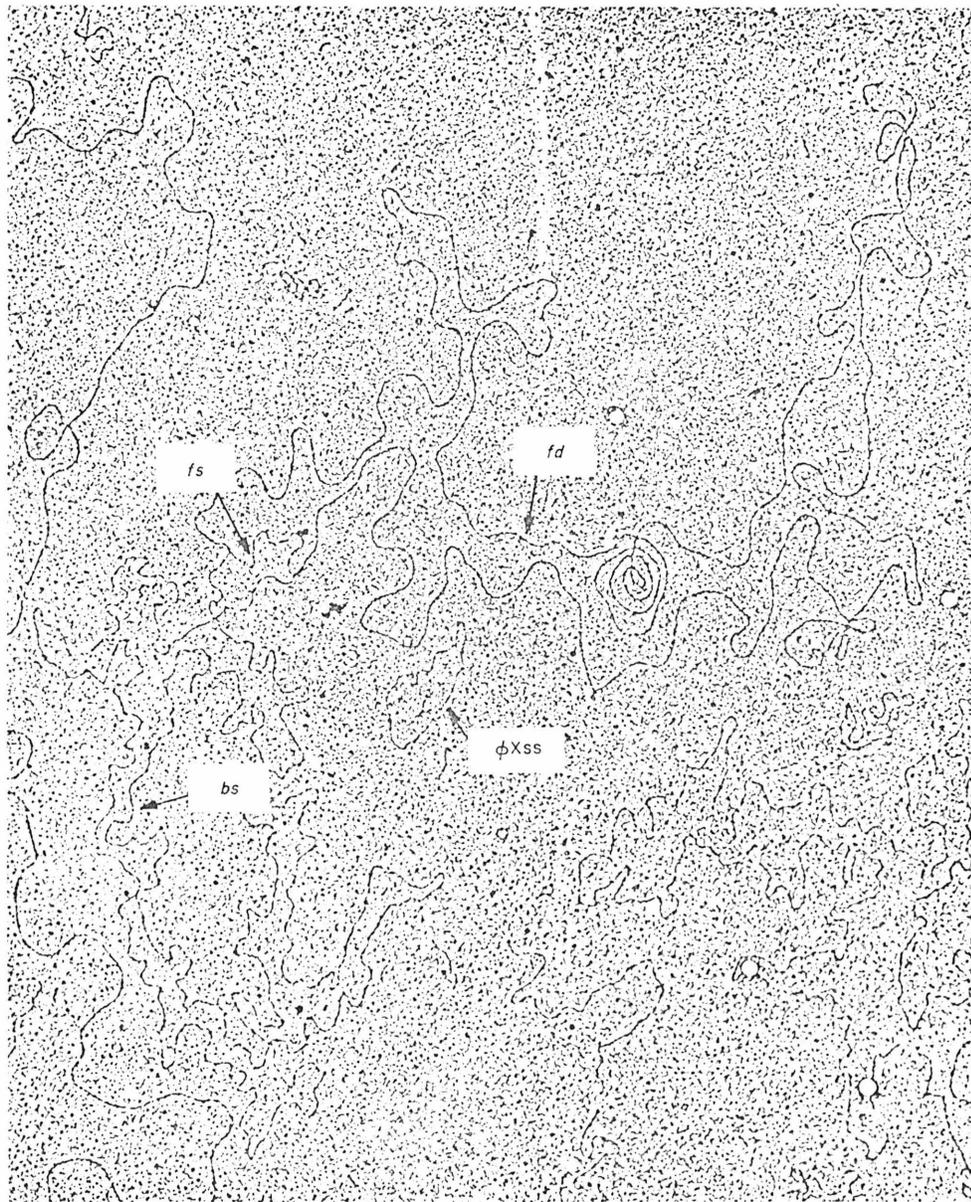


PLATE III. Electron micrograph of an F450/F heteroduplex. Figure 8 is a schematic representation of the structure. The duplex region of F DNA is identified as *fd*. The long single-stranded branch of bacterial DNA is identified as *bs*. The short single-stranded branch of F DNA is labeled *fs*. A small circle of ϕ X single-stranded DNA is identified.

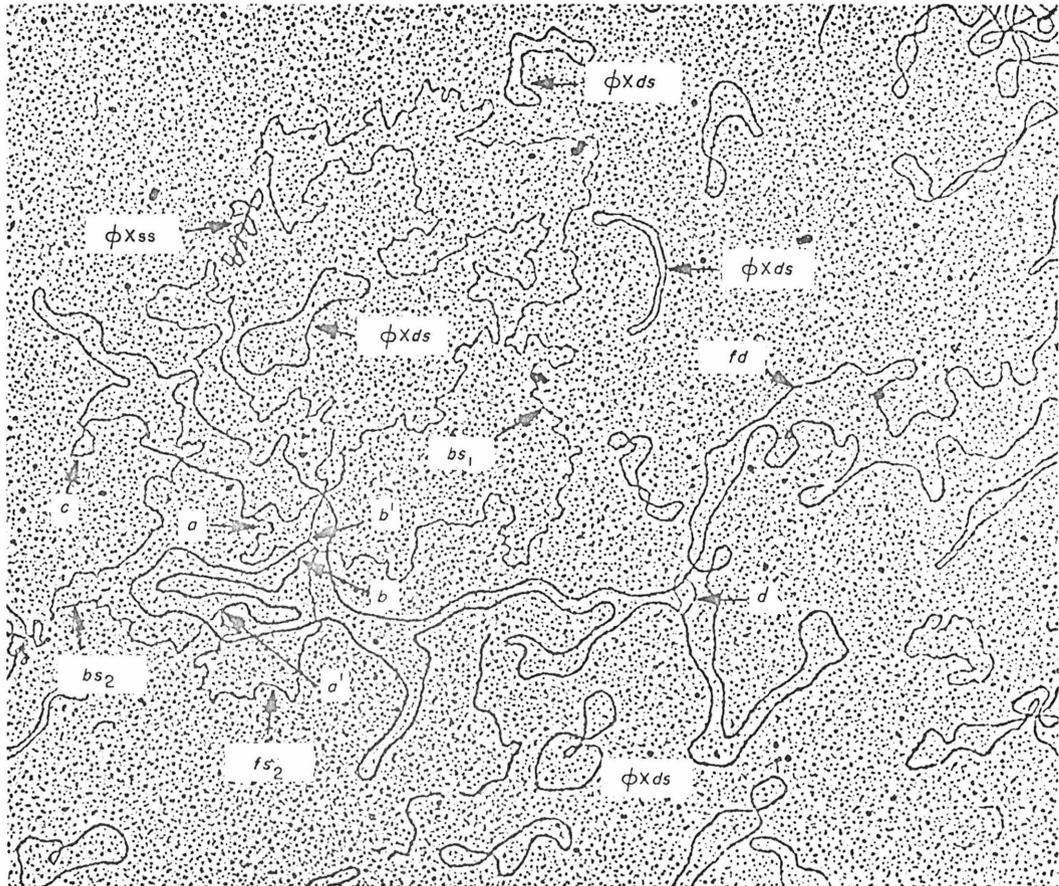


PLATE IV. The *Flac/F₈(N33)* heteroduplex. The structure is shown in Fig. 7(b). The crucial features for identifying the structure are identified by arrows and symbols in the micrograph and in Fig. 7(b), and are discussed in the text. Note that both double-stranded ϕ X-RFI and single-stranded ϕ X are present.

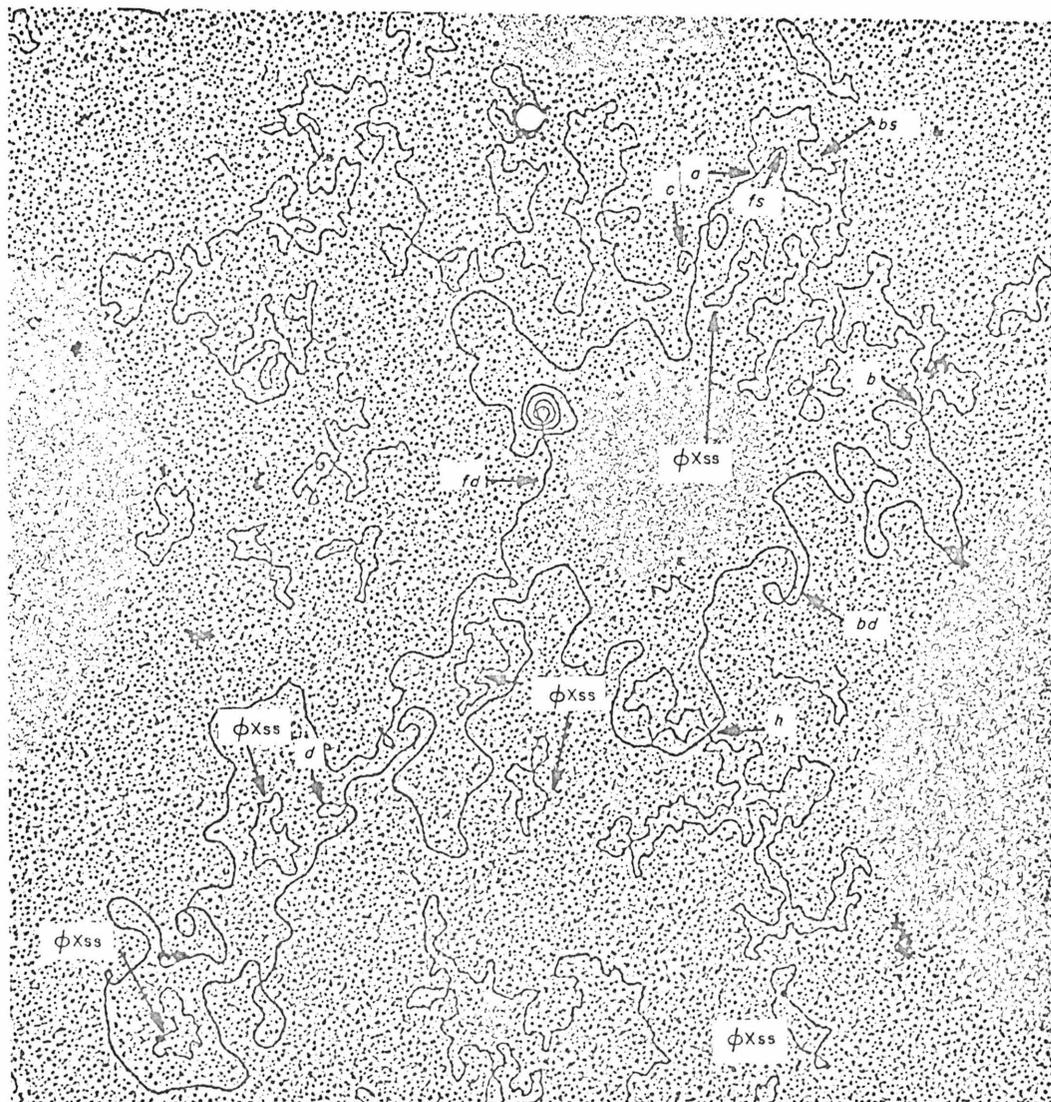


PLATE V. An F450/F₈(N33) heteroduplex. The various symbols can be used to correlate the micrograph with Fig. 9(a). The single-stranded loop *fb*s starts at point *h*. ss, Single stranded.

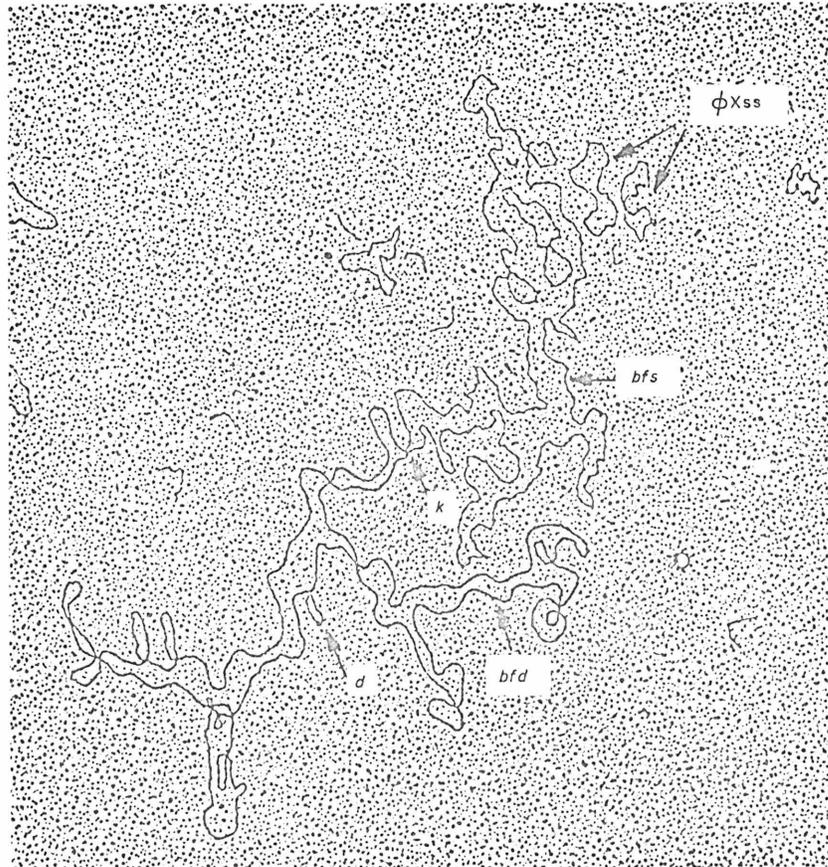


PLATE VI. An $F_8(P6)/F_8(N33)$ heteroduplex. See Fig. 12(b) also. The duplex region is *bfd*. The single-stranded I-D loop, *bfs*, comes out from the duplex DNA, *bfd*, at point *k*. Small loop *d* is also identified. ss, Single stranded.

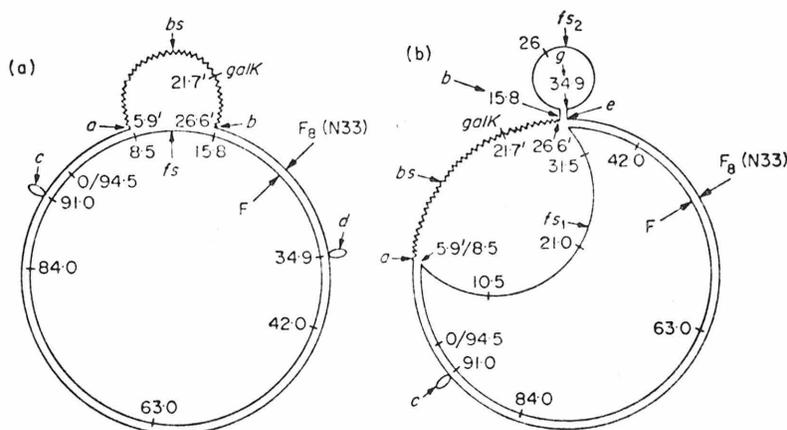


FIG. 3. Heteroduplexes of $F_8(N33)/F$. Note that the F used is from W1485. Solid lines are F DNA; the sawtooth line is bacterial DNA. Unprimed numbers are distances in kilobases from the selected origin on F; primed numbers are distances in kilobases from the selected origin for bacterial DNA. In this, and in other Figures, single-stranded branches of bacterial DNA and of F DNA in a non-homology loop are identified as bs and fs , respectively. Other symbols are explained in the legend to Plate I and/or in the text. (a) As usually seen in complete duplexes without the duplex segment due to the inverted repeat; see Plate I(a). (b) With the duplex segment due to the inverted repeat, as usually seen in incomplete heteroduplexes. See Plate I(b) for a micrograph. The length measurements in kb units and their standard deviations for the several segments in (a) are: fs , 7.3 ± 0.6 ; bs , 20.7 ± 1.1 ; b to d , 19.1 ± 0.8 ; d , 1.4 ± 0.1 ; d to c , 55.7 ± 2.1 ; c , 1.2 ± 0.1 ; c to a , 12.0 ± 0.5 . In this carefully studied case, each feature has been measured at least 10 times.

system for bacterial sequences is simple, generally applicable, and easy to use for precise communication. It is inconvenient in that a given DNA sequence will have different co-ordinates on different plasmids.

The co-ordinate system for F has been arbitrarily defined so that the point a of the episome F450 (see section (d)) is at the origin. It then results that the points a and b of $F_8(N33)$ have F co-ordinates of 8.5 and 15.8 kb, respectively. It is worth noting, for frequent future reference, that the two little I-D loops c and d are at 91.0 and 34.9 kb, respectively.

When nicked $F_8(N33)$ is denatured with NaOH, reneutralized and immediately spread from 50% formamide, before any intermolecular renaturation occurs, each single strand partially self-renatures to give a small duplex segment of length $1.3 (\pm 0.1)$ kb. The molecule is divided into two single-stranded loops by the duplex segment. The shorter and longer loops have lengths of 17.9 and approx. 92 kb, respectively. An electron micrograph is shown in Plate II, and the structure is interpreted in Figure 4. The duplex segment lies between the points e and g in the Plate and Figure. This homology within a single strand is the result of an inverted repeat of a DNA sequence of length 1.3 kb occurring within $F_8(N33)$. The inverted repeat also manifests itself in an alternative structure of the $F_8(N33)/F$ heteroduplex (Plate I(b) and Fig. 3(b)) in which the loop c , at 91 is seen, two single-stranded branches, fs_1 and bs , diverging from the duplex at 8.5 are seen, and the inverted repeat duplex (point e to point g), associated with a single-stranded loop, fs_2 of length

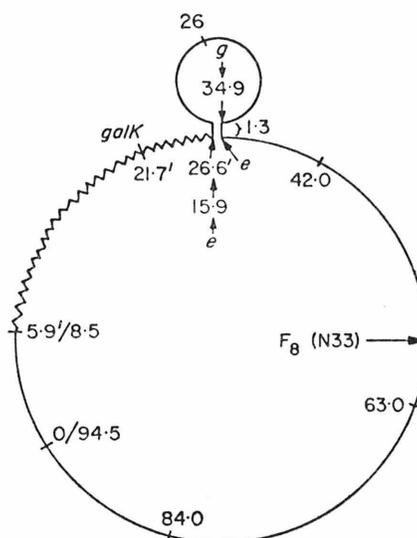


FIG. 4. Structure of a single strand of $F_8(N33)$ DNA in which there is a small duplex segment due to the inverted repeat. See Plate II also. The inverted repeat is between e and g .

17.9 (± 1.1) kb is seen, but the little loop, d , at 34.9 is not present. This result suggests that loop d is one member of the inverted repeat. (The structure in Fig. 3(b) is seen most frequently in partial heteroduplexes between $F_8(N33)$ and a broken F missing the segment from approximately 10 to about 40 so that the duplex segment between 15.8 and 34.8 in Fig. 3(a) cannot form; but complete structures are also seen, as shown in Plate I(b) and Fig. 3(b).)

The suggestion that the I-D loop d at 34.9 is one strand of the inverted repeat is confirmed by the following evidence. As discussed later (Fig. 10), $F_8(N33)$ derives from the parental F_8 in strain W4520. We call this parental episome $F_8(1)$. The structures of the heteroduplexes $F_8(1)/F$ and $F_8(1)/F_8(N33)$ are shown in Figure 5. $F_8(1)$ has neither of the two little I-D loops, c and d . Accordingly, these loops are seen in the $F_8(1)/F_8(N33)$ heteroduplex (Fig. 5(b)). Furthermore, $F_8(1)$ contains an additional piece of bacterial DNA of length 5.9 (± 0.6) kb extending from 0' to 5.9' on the $F_8(1)$ co-ordinate system. When mounted by itself, $F_8(1)$ does *not* show the inverted repeat. Therefore, the little loop, d , at 34.9 present in $F_8(N33)$ but absent in $F_8(1)$ has resulted from an insertion in an inverted order of a DNA sequence close to 15.8 in the F_8 episomes. In all other respects, $F_8(N33)$ and $F_8(1)$ are identical.

The insertion could come from a duplication of either bacterial DNA or F DNA close to the bacterial DNA-F DNA junction at 15.8/26.6' in F_8 episomes. That it most probably comes from F DNA is shown by the structure of the heteroduplex $F_8(N33)/\lambda dg$ LBT2 as shown in Figure 6. This λdg contains *E. coli* bacterial DNA counterclockwise from att^{λ} to a point beyond $galE$ and $galT$, but within $galK$ (it is identified as λdgI^{-} in Kayajanian & Campbell, 1966). Its segment of bacterial DNA extends clockwise from gal to att^{λ} and thus extends further than does the bacterial DNA of $F_8(N33)$. The heteroduplex shows a duplex segment of length 4.9 kb (starting

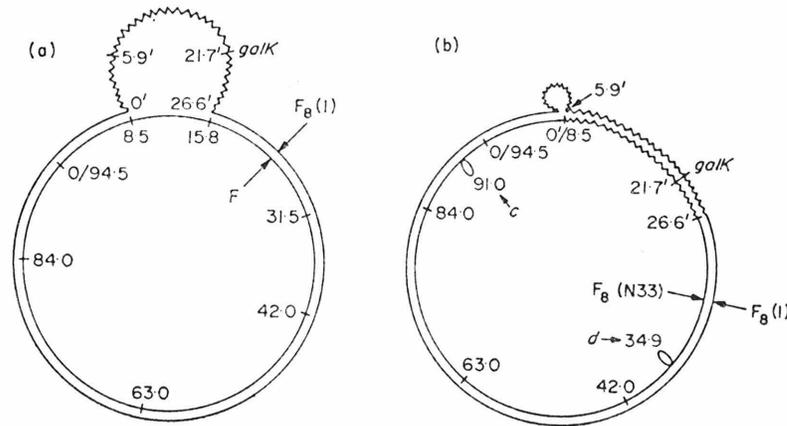


FIG. 5. Heteroduplexes of $F_8(1)/F$ and $F_8(1)/F_8(N33)$. Note that the small I-D loops, *c* and *d*, are present in the latter heteroduplex. Also note that the inference that the sequence of DNA of length 5.9 kb that has been deleted in going from $F_8(1)$ to $F_8(N33)$ is bacterial DNA and that the counter-clockwise excision point of this deletion was exactly at the F-bacterial DNA junction of $F_8(1)$ is based on the equality of the duplex length segment from loop *c* to the 5.9' I-D loop shown in (a) of 12.0 ± 0.2 kb with the distance from loop *c* to the point *a* in Figure 3 of 12.0 ± 0.2 kb. The experimental error in this comparison is ± 0.3 kb, thus it is possible that the illegitimate recombination causing the deletion did not occur exactly at the point 8.5/0' of $F_8(1)$.

at 21.7' in Fig. 6(a)) due to the homology in the *gal* region; it shows the duplex segment, here marked *i* and *i'*, of the inverted repeat, with a single-stranded loop of length 17.8 kb. If the sequence involved in the inverted repeat, labeled *i* in Figure 6(a), were the same bacterial DNA that occurs in *E. coli* K12 at this point, it should be complementary to the sequence of bacterial DNA of λdg LTB2 labeled *a* in Figure 6(a); the structure shown on a reduced scale as Figure 6(b) should then have formed by branch migration (Lee, Davis & Davidson, 1969), since it is thermodynamically a more stable structure for the same number of base pairs as in Figure 6(a); but the structure shown in Figure 6(a) is observed, and that in Figure 6(b) is not. Thus, the reverse duplication of $F_8(N33)$ (and of its parental episome, $F_8(JE3100)$, see section (i)) was probably formed by duplication of the pieces of F DNA from 15 to 16.2 and its insertion in reverse order into the F sequence of F_8 at point 34.9.

Among the F_8 episomes studied, the loop at 91 occurs only in $F_8(N33)$. $F_8(N33)$ was produced by nitrosoguanidine mutagenesis of $F_8(JE3100)$ (see Fig. 10) and selection for a transfer-defective mutant; it was found to be defective in the *traA* cistron (Ohtsubo *et al.*, 1970). Careful length measurements on heteroduplexes of $F_8(N33)$ and F with ColV, which are partially homologous in this region (Sharp, Cohen & Davidson, 1973) show that the loop at 91 is most probably a small insertion. It occurs within a region of F DNA containing the *traA* cistron (see section (j) below); it is therefore presumably the cause of the defect in *traA* cistron of $F_8(N33)$. The loop appears to contain an *E. coli* transfer RNA gene (Wu & Davidson, personal communication); thus it is probably an insertion of bacterial DNA.

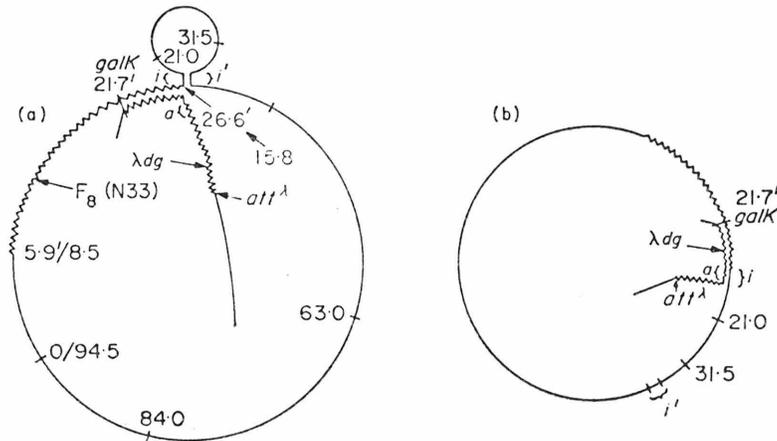


FIG. 6. Heteroduplexes of $F_8(N33)$ with $\lambda dgLBT2$. In the λdg , λ DNA is shown as a straight line and bacterial DNA as a sawtooth. Structure (a), in which the duplex segment of $F_8(N33)$ consisting of complementary strands i and i' due to the inverted repeat occurs, is seen; structure (b) in which segment i forms a duplex with segment a of the λdg is not seen. This result shows that i is F DNA, not the bacterial DNA sequence that occurs at this point in *E. coli* K12.

(c) *Flac*

The structure of the *Flac*/F heteroduplex is schematically shown in Figure 7(a). *Flac* is missing a segment of F DNA of length 3.0 ± 0.4 kb (co-ordinate 0 to 3.0 in Fig. 7(a)), suggesting that it was formed by a type I excision. The bacterial DNA has a length of 51.7 ± 1.9 kb.

(d) *F450*

This episome, harbored in strain B583, carries alleles for *gal*⁺, *att*^λ, *bio*⁺, *uvrB*⁺ (Freifelder, Folkmanis & Kirschner, 1971) in a host with a chromosome deleted from a point counterclockwise of *galE* to a point clockwise of *chlA* (see footnote e of Table 1).

An electron micrograph of the *F450*/F heteroduplex is shown in Plate III; the structure is shown in Figure 8. Again, the structure is that expected for an F' formed by a type I excision process. The missing piece of F DNA has a length of 3.0 ± 0.4 kb, the bacterial DNA is of length 80.0 ± 1.5 kb.

(e) Relative co-ordinates on F

F450, *Flac* and the F_8 episomes have a segment of F DNA missing of length 3.0, 3.0 and 7.3 kb, respectively. What are the relative map positions of these several segments? Where on F do the junctions of bacterial DNA and F DNA occur in the several episomes? As already remarked, because of the unusual structural features of $F_8(N33)$, heteroduplexes between it and the other episomes can be used to answer these questions. The two small I-D loops, *c* and *d*, of $F_8(N33)$ at positions 91 and 34.9, respectively, of F can be used as reference points to map the deletions of F DNA and the junctions of bacterial with F DNA in other episomes. As will be seen, the inverted repeat of $F_8(N33)$ DNA is also used in these analyses.

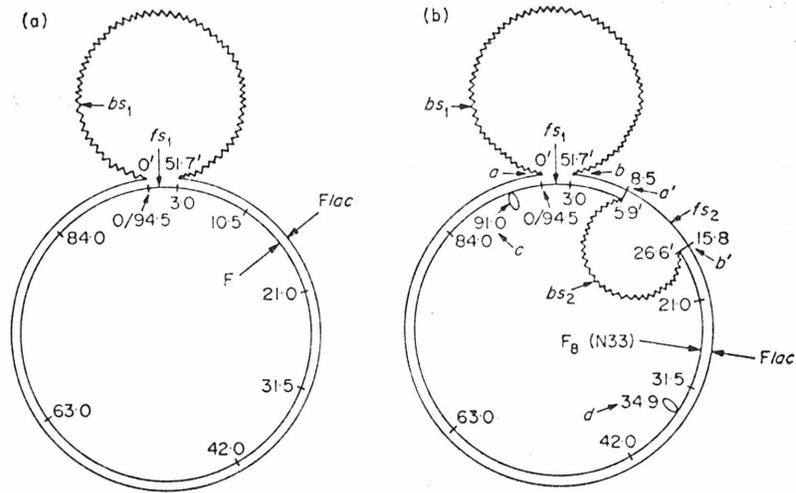


FIG. 7. Heteroduplexes involving *Flac*. Plate IV is a micrograph corresponding to Fig. 7(b).

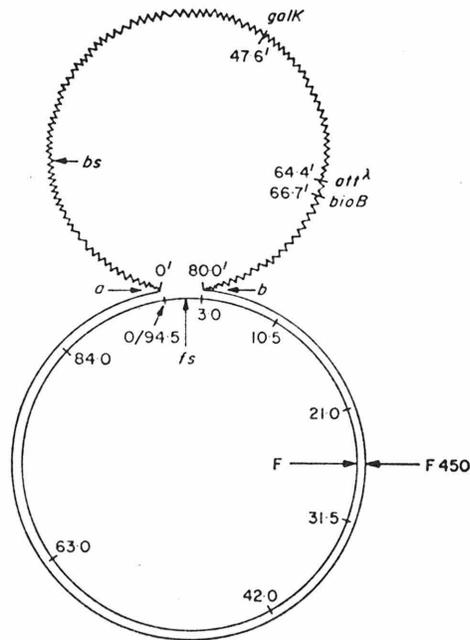


FIG. 8. Heteroduplex structure of F450/F. An electron micrograph is shown in Plate III.

The structure of the heteroduplex *Flac*/F₈(N33) is shown in Plate IV and Figure 7(b). We now see two large substitution loops separated by a duplex segment of length 5.5 kb. One loop with single-stranded branches, denoted by *fs*₁ and *bs*₁, has the same dimensions as the *Flac*/F substitution loop. It is 3.5 kb from one of the small insertion

loops. The other large substitution loop, with branches fs_2 and bs_2 , is like the $F_8(N33)/F$ substitution loop; it is 19.1 kb from the other small insertion loop. These quantitative considerations show that the 3.0 and 7.3 kb pieces of F deleted in *Flac* and $F_8(N33)$ are separated by a sequence of F of length 5.5 kb; the sequence missing in *Flac* is counterclockwise to the sequence missing in F_8 .

In a similar way as for *Flac*, the relative location of the 3.0 kb segment of F deleted in F450 was deduced from a study of the F450/ $F_8(N33)$ heteroduplex. This case is more complicated than that of *Flac*/ $F_8(N33)$ because the bacterial sequences present in $F_8(N33)$ are also present in F450. Two of the structures seen are shown in Figure 9(a) and (b). The more stable structure is shown in Figure 9(a); a micrograph is shown in Plate V. Again, one small I-D loop is 3.5 kb from the beginning of a large substitution loop with branches fs and bs . The other small I-D loop is again 19.1 kb from the large I-D loop, lbs , in Figure 9(a). These small loops are therefore c and d , respectively. The structure shown in Figure 9(b) in which the reverse repeat manifests itself is seen most frequently in partial structures, where F DNA from about 15.8 to 36 is missing; it is, we believe, less stable than the structure shown in Figure 9(a) in complete structures. Many partial structures of the type suggested in Figure 9(a) are seen in which duplex formation between the two segments connected by a dotted line occurs. (But this duplex formation does not occur in the sample shown in Plate V.) These mating segments are complementary segments of F, between 3.0 and 8.5 kb, which are, respectively, counterclockwise and clockwise from the bacterial DNA in $F_8(N33)$ and F450. In a complete structure, it is sterically difficult for them to pair. Nevertheless, such complete structures do occur with low frequency, but they are usually so tangled that it is difficult to trace their topology. All of the quantitative length measurements lead to the structural assignment shown in Figure 9(a) and to the numerical assignments for relative positions of F and bacterial sequences given in the several Figures.

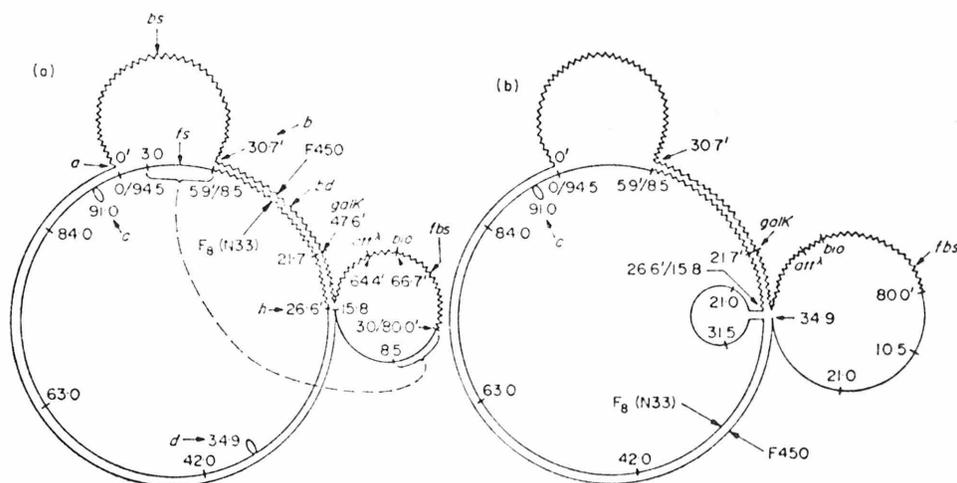


FIG. 9. Structures seen for F450/ $F_8(N33)$ heteroduplexes. Structure (b) in which the short duplex segment due to the inverted repeat has formed is most frequently seen in heteroduplexes between incomplete strands. The homologous regions 3.0 to 8.5 of F in the two strands sometimes pair in structure (a) as suggested by the dotted line. Plate V is a micrograph of the structure depicted in Fig. 9(a).

In particular, if the 3.0 kb piece of F missing in F450 is assigned co-ordinates 0 to 3.0, the piece of F missing in the F_8 episomes has co-ordinates 8.5 to 15.8 kb. *The important result is that to within the experimental error of our measurements the same sequence of F (0 to 3.0 kb in our numbering system) is missing in Flac and F450, although these two episomes were derived from different Hfr's with F inserted at quite different points in the E. coli K12 chromosome.* As will be reported later in section (k), the same segment is missing in F_{2gal} , an episome which was derived independently from the same Hfr (Hfr 3) as was F450 (Liedke-Kulke & Kaiser, 1967).

(f) *Bacterial sequences in F450 and $F_8(N33)$*

The data presented in the previous section and the data on the $F_8(N33)/\lambda dg$ heteroduplex (Fig. 6) lead to the assignments of bacterial sequences shown in Figure 9. These interpretations were confirmed and additional mapping data obtained by making a heteroduplex of F450(λ) (λ inserted into F450) and $F_8(N33)$ and a di-heteroduplex F450(λ)/ $F_8(N33)/\lambda b2b5c$. In the heteroduplex F450(λ)/ $F_8(N33)$ the size of the substitution loop is unchanged but the I-D loop *lbs* of Figure 9 which contains *att^λ* becomes larger by the size of λ ; this confirms the identification. The position of the *att^λ* junction (as shown in Figs 8 and 9) was measured in the di-heteroduplex F450(λ)/ $F_8(F33)/\lambda b2b5c$ (Sharp, Hsu & Davidson, 1972).

The observed structures, with the data already presented for the $F_8(N33)/\lambda dg$ heteroduplex, not only fix the positions of the *gal* genes but show that *gal* DNA has the same polarity relative to the F strands in F450 and in $F_8(N33)$. The *bio* position shown in our map is 2.3 kb clockwise from *att^λ* and is the position of the $\lambda bio16A$ cut as physically mapped by Hradecna & Szybalski (1969). (The transducing phage $\lambda bio16A$ carries bacterial DNA starting at *att^λ* and ending within *bioB*.)

(g) F_{1s}

$F_1 gal att^{\lambda} bio$ is an F' isolated by Jacob & Wollman (1961). Signer, Beckwith & Brenner (1965) report that this episome, as sent to them by Wollman, can carry the *supE* amber-suppressor and *supB* ochre-suppressor alleles. A bacterial strain carrying the episome was sent to us by E. Signer. We refer to it as F_{1s} to allow for the possibility that it is not identical to the original Paris isolate. As noted in footnote e of Table 1, F450 is also a descendant of the original Paris $F_1 gal att^{\lambda} bio$.

We find that F_{1s} has a molecular length of about 227 (± 6) kb; thus it is 55 kb larger than F450. Studies of the heteroduplexes F_{1s}/F and $F_{1s}/F_8(N33)$ show that the same piece of F (0 to 3.0 kb) is deleted in F_{1s} as in F450.

Further studies have shown that F_{1s} and F450 contain the same bacterial DNA from 30.7' to 80' of F450 (thus 85.7' to 136' kb of F_{1s}), that is, in the *nadA*, *aroG* (see later), *gal*, *att^λ* and *bio* regions, and from 0' to about 6.8' of F450 which is homologous with 0' to 6.8' of F_{1s} . The bacterial DNA sequences between 6.8' and 30.7' of F450 are not homologous to the sequences from 6.8' to 86' of F_{1s} . Further studies of the physical and genetic structure of F_{1s} and F450 will be reported later. These studies are consistent with the hypothesis that F_{1s} and F450 have both been derived by deletion of some bacterial DNA from a common $F_1 gal att^{\lambda} bio$ parent.

(h) *F(W1485)*, *F(YS40)* and *F(W6)*

The pedigrees of F leading to the Hfr's from which F_8 , F450 and F_{1s} , and *Flac* were derived are discussed in the footnotes to Table 1. The Hfr's were probably all derived from strain 58-161 (F^+) which was derived in several steps from the original K12 (F^+). W1485 (F^+) was directly derived from the original K12 (F^+). The heteroduplex studies reported here have been done with the F from W1485. Our conclusions about the processes leading to the several F' episomes would be invalid if the episomes were derived from different F factors. YS40 (F^+) is a strain received from D. R. Helinski, containing F^+ transferred from W1485. F from W1485 and F from YS40 have the same molecular length (Table 3); heteroduplexes of either with F450 give the same structure. A heteroduplex of a mixture of the two F factors looked just like a homoduplex of either with no non-homology regions. The two F factors are therefore the same by the heteroduplex criterion.

W6(F^+) is a spontaneous *bio*⁺ revertant of 58-161(F^+). The F_8 episomes were derived in several steps from W6. Closed circular DNA molecules were prepared from a culture grown from an old stab of W6 kindly supplied by E. Lederberg. All native duplex molecules had the same length within experimental error. Duplex molecules observed after denaturation and renaturation of the sample showed no large deletions, such as the F deletions in F450 or F_8 . A small fraction, perhaps 5 to 10%, of the duplexes showed an I-D loop with a size of about 1.2 kb. After denaturation, reneutralization and rapid spreading, a comparable percentage of the molecules contained an inverted repeat of length 1.2 kb. In most cases, the inverted complementary sequences were separated by a single-stranded loop of length 10.4 kb; individual cases of spacings of 15 and 20 kb were seen. Heteroduplexes with $F_8(N33)$ showed *only* the structure seen in Figure 3 in the region clockwise from loop *c* to loop *d*. The sample of F(W6) is therefore somewhat heterogeneous. However, it appears that F(W6) and F(W1485) are the same as regards all features which are important for the interpretation of the structure of the several F-primes studied here, in particular the sequence between 0 and 3 kb.

We therefore believe that most probably the episomes studied were all derived from a common F by the processes indicated in Figure 1.

(i) F_8 episomes

The treatments leading to the several F_8 episomes are shown in Figure 10. Deletion mapping of the F_8 -primes prepared by P1 transduction will be discussed in the next section.

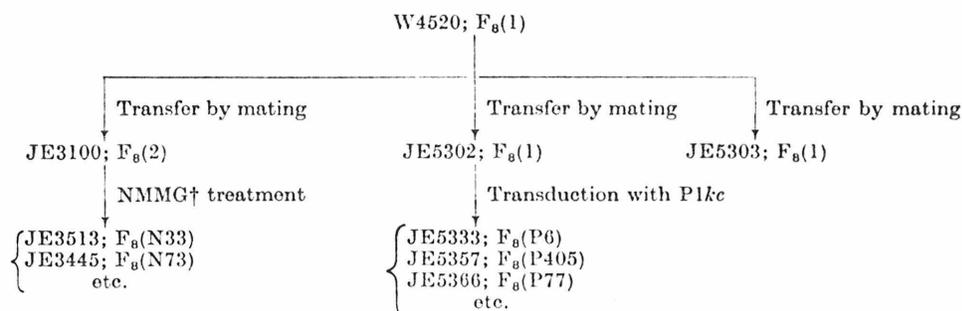


FIG. 10. Pedigree of the several F_8 episomes. W4520 containing F_8 was isolated by Hirota & Sneath (1961); all other transfers are by Ohtsubo *et al.* (1970) and Ohtsubo (1970).

† NMMG, nitrosoguanidine mutagenesis.

The parent of all the F'_8 -primers studied here is F'_8 (W4520), isolated by Hirota & Sneath (1961). It is a very fertile sex factor and is readily cured by acridine orange. F'_8 (JE5303) was prepared by transfer of F'_8 (W4520) into the recipient strain N23-53. These two F'_8 episomes give identical heteroduplexes with F. We shall refer to these two apparently identical episomes as F'_8 (1). F'_8 (1) has the following structural features. (a) In heteroduplexes with F (Fig. 5), it does not show the two I-D loops, *c* and *d*, that F'_8 (N33) shows. Its piece of bacterial DNA is $5.9(\pm 0.4)$ kb larger than that of F'_8 (N33). (b) Accordingly, in a heteroduplex, F'_8 (1)/ F'_8 (N33) (Fig. 5(b)), the two loops, *c* and *d*, are seen. Furthermore, a single-stranded deletion loop of length 5.9 kb occurs at the junction at 0'/8.5 between F DNA and bacterial DNA. This is the extra piece of non-F DNA in F'_8 (1). We refer to it as the F'_8 (1) (0'-5.9') segment. (c) Note that, because of the way we define the co-ordinate systems, the same *galK* sequence has co-ordinates 21.7' and 47.6' in F'_8 and F450, respectively. The F'_8 (1) (0'-5.9') segment has its right terminus 15.8 kb to the left of *galK*. The F450 (25.9'-31.8') segment also has its right terminus 15.8 kb to the left of *galK*. But we have observed that in an F450/ F'_8 (1) heteroduplex, these two segments are not homologous. However, in the F'_{1s} / F'_8 (1) heteroduplex, homology is seen in this segment of DNA, suggesting that as discussed in section (g), F450 is derived from the "original" F'_1 *gal att^A bio* by a deletion in this region. (d) In self-renatured DNA from an uncloned sample from the original stab of W4520, some small I-D loops are seen—indicating some heterogeneity.

F'_8 (JE3100) (which we call F'_8 (2)) was prepared by Ohtsubo by mating of W4520 (F^+) with JE2571. It was then used, as indicated in Figure 10, to prepare F'_8 (N33) and various other transfer-defective mutants by nitrosoguanidine mutagenesis and M12 selection. The stab of JE3100 available for investigation was rather old. The closed circular DNA obtained was somewhat heterogeneous in that some small I-D loops were seen in a self-renatured sample. However, the main component resembled F'_8 (N33) in that it did not contain the 0' to 5.9' sequence of F'_8 (1), and it does have the loop, *d*, at 34.9 and the corresponding inverted repeat. Thus, the loss of the F'_8 (1) (0'-5.9') sequence and the acquisition of the inverted repeat occurred in some unknown way during the processes leading to the isolation of JE3100. F'_8 (2) resembles the parental F'_8 (W4520) in the several sex factor phenotypes for which it has been tested (Ohtsubo *et al.*, 1970). (We think, from our incidental observations, that inverted repeats accumulate spontaneously upon storage of certain episome-bearing strains.)

It was found that the episome F'_8 (N73) which was derived from F'_8 (2) by nitrosoguanidine mutagenesis and which is defective in the *traB* cistron has the same physical structure, by heteroduplex studies, as does F'_8 (2). Therefore, the mutation in the *B* cistron is probably a point mutation, or at least one involving less than 50 base pairs and hence not observable by heteroduplex studies.

As shown in the Figure, JE5302 (F'_8 ⁺) was the parental strain for preparation of the deletion mutants discussed below. It was produced by mating W4520 (F'_8 ⁺) with W3623 (*str*^r). Plasmids and episomes introduced into W3623 are rather unstable and are frequently lost during cultivation or storage. In fact, the stab of JE5302 (F'_8 ⁺) available for investigation by us had lost the *gal*⁺ character and presumably the F'_8 episome. However, F'_8 (P77) which had been prepared previously by transduction from F'_8 (JE5302) has exactly the same heteroduplex structure and the same genetic properties as F'_8 (1). Therefore, we believe that F'_8 (JE5302) had the same structure as F'_8 (1).

(j) *Deletion mutants of F₈*

Ohtsubo (1970) has previously prepared a number of transfer-defective deleted F₈'s by P1*kc* transduction (Fig. 10). The deletions were genetically mapped with respect to *gal* genes and to certain transfer cistrons. We have now studied the physical structure of these deleted F₈'s. Further genetic studies to map the bacterial markers *nadA* and *aroG* on F₈ have also been carried out.

(i) *Genetic studies*

The host was a bacterial strain PL225 carrying the mutations *nad* and *gal* which may have occurred in a single mutational event (Mizuuchi & Fukasawa, 1969). These mutations in PL225 are not revertable, so PL225 is believed to have a deletion. PL225 cannot utilize quinolinate, so it is not *nadB*; quinolinate phosphoribosyl transferase in PL225 is normal so it is not *nadC* (Fukasawa, personal communication); therefore, the *nad* character in PL225 may be *nadA* which is closely linked to *gal* (Taylor & Trotter, 1967). We observed that PL225 did not grow on a minimal plate containing 25 µg tryptophan/ml. and 25 µg of tyrosine/ml. in addition to 2 µg of niacin/ml. and the original requirement of 2 µg of thiamine/ml. This shows that PL225 has the *aroG* mutation. The *aroG* marker is also known to be closely linked to *gal* (Wallace & Pittard, 1967; Adhya, Cleary & Campbell, 1968). These observations can be accounted for if PL225 is deleted in a region of the chromosome spanning *nadA*, *aroG* and *gal*.

PL225 F₈⁺ was prepared by a cross between JE5303, F₈⁺ *trp arg*, and PL225, F⁻ *nadA thi gal*, by selecting *gal*⁺ *arg*⁺ *trp*⁺ on a selective plate containing 2 µg of thiamine/ml. and 2 µg of niacin/ml. with 1% of galactose as the only carbon source. This PL225 F₈⁺ strain could grow on a minimal plate containing tryptophan, tyrosine, niacin and thiamine, as described above, and on a minimal plate containing only added thiamine (2 µg/ml.). Thus, the F₈ episome carries *nadA*⁺ and *aroG*⁺. The mating tests were readily checked in transfer experiments by cross-streaking an F₈ donor against

TABLE 4
Genetic analysis of F₈ deletion mutants

Mutants	Deletions identified	<i>nadA</i>	<i>aroG</i>
F ₈ (P44), F ₈ (P101), F ₈ (P106) F ₈ (P107), F ₈ (P144), F ₈ (P124) F ₈ (P133), F ₈ (P215), F ₈ (P229)	Δ(<i>galK</i>)	—	—
F ₈ (P11), F ₈ (P54)	Δ(<i>traA</i>)†	+	+
F ₈ (P43), F ₈ (P219), F ₈ (P321)	Δ(<i>traA-traB</i>)	+	+
F ₈ (P19), F ₈ (P432)	Δ(<i>traA-traB-traC</i>)	+	+
F ₈ (P17)	Δ(<i>traA-traB-traC</i>)	—	+
F ₈ (P322)	Δ(<i>traA-traB-traC</i>)	—	—
F ₈ (P6)	Δ(<i>traA-traB-traC-traD-traE</i>)	—	+
F ₈ (P376)	Δ(<i>traA-traB-traC-traD-traE</i>)	—	—
F ₈ (P405)	Δ(<i>galK-traA</i>)	—	—
F ₈ (P82)	Not detected	—	—
F ₈ (P77), F ₈ (P78), F ₈ (P81)	Not detected	+	+
F ₈ (1)	Wild type	+	+

† See the discussion of the notation for the *tra* cistrons in the legend to Fig. 11.

recipient PL225 on the selective plates described above. Several of the deletion mutants of F'_8 were tested for complementation with *nadA* and *aroG* by this cross-streaking method.

The deletion mutants of F'_8 studied here, some of which are transfer defective, were available in strains carrying the R factor, R100-1, thus rectifying the defective fertility of these F'_8 's (Ohtsubo *et al.*, 1970). Table 4 lists the results of complementation tests for *nadA*⁺ and *aroG*⁺ with the deleted F'_8 's. Four transduction derivatives of F'_8 which carried no detectable *tra* and/or *galK* deletions were also tested. One of them lacked both *nadA* and *aroG*, but the others carried all the chromosomal genes tested. The results of all of these genetic mapping experiments are shown in Figure 11. By the overlapping deletion method, the order of the genes on F'_8 is deduced as *gal*-(*O-E-T-K*)-*aroG-nadA-traA-traB-traC*-(*traD-traE*)-*traF*. The order of the bacterial

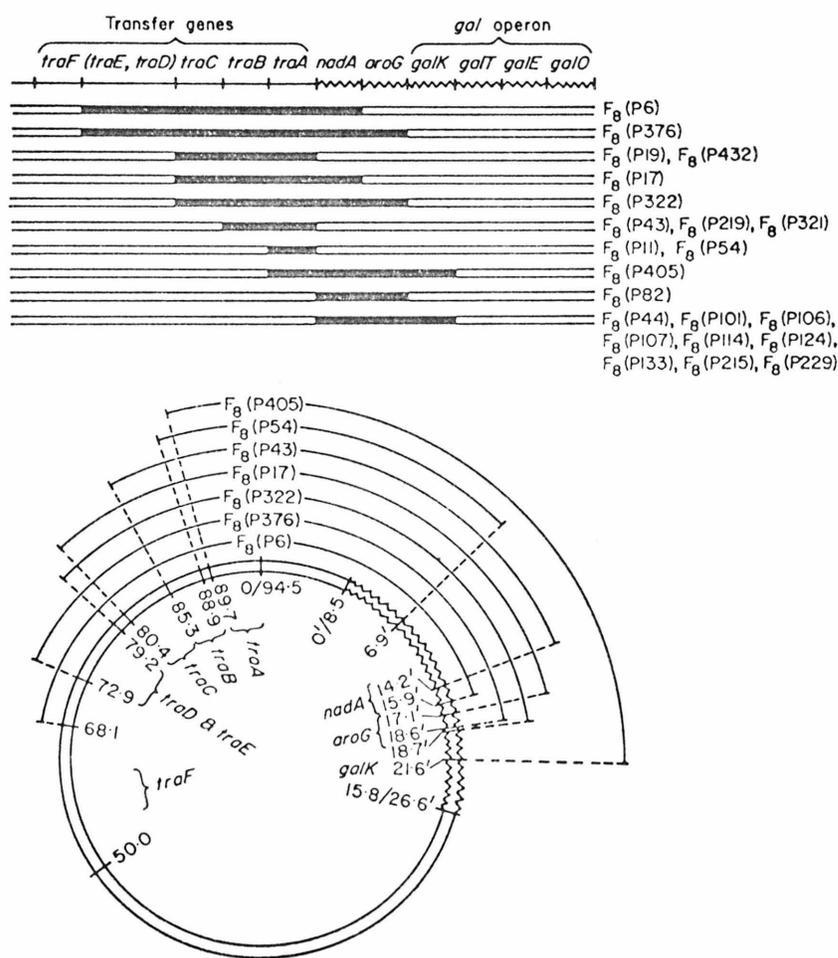


FIG. 11 Results of genetic (upper diagram) and physical (lower diagram) mapping of the F'_8 episomes prepared from $F'_8(1)$ by P1 λ transduction by Ohtsubo. It should be noted that the *tra* cistron notations used here, as introduced by Ohtsubo *et al.* (1970), differ from those used by Willetts & Achtman (1972) and by Ihler, Achtman & Willetts (1972). For further discussion, see text.

chromosomal genes is quite consistent with the *E. coli* K12 map arranged by Taylor (1970).

(ii) *Physical studies*

Physical structures were studied by forming heteroduplexes with $F_8(N33)$ and F . The heteroduplexes with $F_8(N33)$ have simple, readily interpretable structures corresponding to a simple deletion in each case. Plate VI shows an example of the $F_8(P6)/F_8(N33)$ heteroduplex with the structural features interpreted in Figure 12(b). It should be noted that $F_8(N33)$ is itself deleted of about 5.9 kb of bacterial DNA relative to the parental $F_8(1)$ from which the deletion mutants of F_8 were derived. Thus, the actual deletion size in the F_8 mutants is 5.9 kb longer than that seen in the heteroduplexes. The insertion loop *d* at 34.9 serves as a fixed point to confirm the structural assignments. A characteristic feature in all the heteroduplexes of $F_8(N33)$ with the F_8 mutants, all of which are genetically analyzed to have deletions in *traA*, is that the insertion loop, *c*, at 91 is not observed. This is consistent with the interpretation suggested above that this insertion loop is in the *traA* cistron.

Heteroduplexes between these F_8 mutants and F were also studied. One such structure is shown in Figure 12(a). The structure of a deleted F_8 can be inferred from its heteroduplex with F . The F arm of the substitution loop increases from 7.3 kb by an amount equal to the additional amount of F deleted in the F_8 mutant; the length of the F_8 arm can then be used to calculate the amount of bacterial DNA deleted. Most of our measurements were in fact made in such F_8/F heteroduplexes. A summary of the structural features of all of these deletion mutants is presented in Figure 11. Correlation of the genetic data and the physical data fixes the *galK* site at $21.6' \pm 1.0$ kb in $F_8(1)$ in agreement with the mapping of the $\lambda dg/F_8(N33)$ heteroduplex reported above. The correlation of physical and genetic data map *nadA* as 6 kb counterclockwise from *galK*. *aroG* is in between these two markers. Note that the episome $F_8(P17)$ was originally classified as a mutant deleted in the region from *traA* to *traD*, but more recently we have found in complementation tests that the F markers deleted extend only from *traA* to *traC*. The physical mapping confirms this.

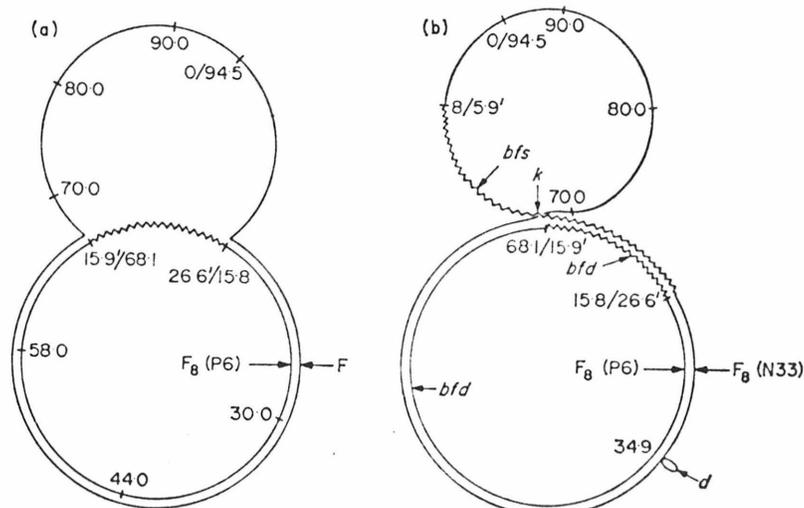


FIG. 12. Structure of $F_8(P6)$ heteroduplexes. Figure 12(b) correlates with Plate VI.

Thus, no mutants are available with which to fix the order of the *traE* and *traD* markers. Although no mutants for mapping the *traF* marker are available, we surmise that it is located in the 50 to 65 kb region of F as discussed below.

(k) *A deleted F*

F_{2gal} is a *gal*-bearing episome derived from Hfr3 by Wollman (Liedke-Kulke & Kaiser, 1967). Its properties have been described by Herbert & Guest (1968).

A slant of strain W3101 $F' gal$, believed to carry F_{2gal} , was kindly sent to us by Professor A. D. Kaiser. Typical colonies were *gal*⁺ and male. Closed circular DNA was extracted from cultures grown in Tryptone broth; the resulting DNA was the deleted F, $F\Delta$ (0-14.5) described below.

Male, *gal*⁺ colonies of W3101 $F' gal$ were mated with N23-53 $F^- gal recA str^r$. No $F^+ gal^+ str^r$ F-ductants or recombinants were found, but many $F^+ str^r gal$ colonies were produced. (In these experiments, F^+ character is defined as sensitivity to the male specific phage, M13.) Closed circular DNA was extracted from one of these. The structure is shown in Figure 13. The remarkable fact is that the episome is a deleted F; the deletion runs from 0 to 14.5 kb. *The deletion therefore starts at the same point on F, within experimental error, as does one of the two junctions between F DNA and bacterial DNA in F450, F₁₈, and F_{lac}.* Note that $F\Delta$ (0-14.5) contains the genes essential for fertility. The identical deleted F was extracted from a slant of F_{2gal} in PA106 kindly sent to us by Drs P. Broda and J. Gross from the collection at the Medical Research Council Laboratory in Edinburgh. Bacteria cultured from this slant were all *gal*⁻, but sensitive to male specific phage.

We presume that, in these strains, F_{2gal} readily reinserts *gal* into the chromosome or loses its bacterial DNA. The points at 0 and 14.5 are evidently hot spots for the recombination event involved in this segregation.

Efforts to obtain an episome that can transfer *gal* from " F_{2gal} "-bearing strains are continuing. We have recently succeeded in isolating F_{2gal} episomes carrying bacterial genes. They are deleted in the 0 to 3 kb region of F. Further studies will be reported later.

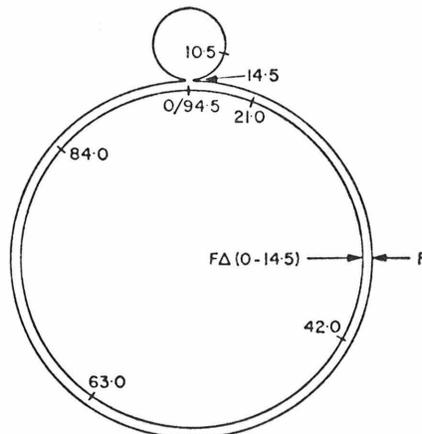


FIG. 13. Structure of the deleted F, $F\Delta(0-14.5)$. The deletion was mapped by constructing a heteroduplex with $F_6(N33)$.

4. Further Discussion

A representation of the structure of F based on the physical data obtained in our work and previous genetic studies is given in Figure 14. The specific positions of the several mapped *tra* cistrons of F are shown in the F_8 map in Figure 11. The *traA* cistrons map by deletion between 87 and 8·5. We assign it to the region between 87 and 94·5 because of the fertility of the $F\Delta$ (0-14·5) episome, which is deleted between 0 and 14 and the fertility of F_{450} and F_{lac} which are deleted between 0 and 3·0. All the other transfer cistrons that have been mapped lie within the region 65 to 94·5.

It is known that there is functional homology between F factors, and certain R factors and colicin plasmids, classified as *fi*⁺. This functional homology includes the ability to complement defects in fertility. Sharp, Cohen & Davidson (1973) have studied the heteroduplexes of F with several such R factors and with the ColV-K94 plasmid. They find that there is a region of DNA homology from around 50 to 94·5 kb on F with small amounts of non-homology consisting of substitution or I-D loops. This is strong evidence that the *tra* functions are contained between 50 and 94·5 kb region of F.

Only five *tra* cistrons have been mapped in a region from 65 to 94·5 kb. The *traF* cistron is known to lie outside of the 65 to 94·5 kb region. Since it can be complemented by an R factor, it probably lies between 50 and 65. Two other *tra* cistrons have been identified by Ohtsubo *et al.* (1970) but their genetic positions are unmapped. Recently, Willetts & Achtman (1972) have identified eleven cistrons involved in transfer of F, and T. Miki (personal communication) has identified at least 14. Willetts & Achtman (1972) and T. Miki (personal communication) report studies of the relations between their mutations and those of Ohtsubo *et al.* (1970). An independent study of the order of the *tra* genes has been reported by Ihler *et al.* (1972). (For a general review, see

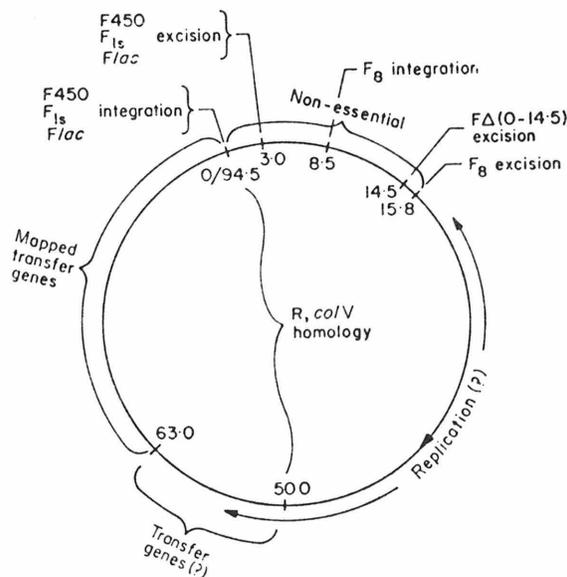


FIG. 14. Summary of the structure of F based on physical and genetic studies.

Willetts (1972).) Since the segment from 50 to 94.5 kb has enough DNA for about 40 typical genes, there is ample room in this region for all of the known *tra* markers.

As discussed below, the results with F450, F_{1s}, *Flac*, and the F₈ episomes indicate that the region 0 to 8.5 can be used for insertion into the bacterial chromosome.

In F450, F_{1s} and *Flac* the region 0 to 3 kb is deleted; in F₈, the region 8.5 to 15.8 is deleted. In FΔ (0-14.5) 0 to 14.5 is deleted. Thus, the DNA of F in the region 0 to 15.8 is not essential and has no known function. The F₈ episomes deleted in the region of F between 94.5 and 65 (Fig. 11) are defective in transfer but are capable of autonomous replication. They can be complemented for transfer by certain R factors; thus they probably contain the structural element for the origin of transfer. Thus the genes for autonomous replication and the structural element for the origin of transfer probably all lie between 15.8 and 65. The direction of transfer is as indicated in Figure 14.

The heteroduplex structures for F450, F_{1s}, F₈ and *Flac* are consistent with insertion of F to form Hfr's as indicated in Figure 15 and excision by a type I process at the points indicated. In this interpretation, the insertion of F at the same point (0/94.5) occurred at two different places on the bacterial chromosome for Hfr2 and Hfr3; in both cases, excision occurred at 3.0 kb. (Alternative interpretations are possible. For example, the F used by Jacob & Wollman (1961) to make Hfr3 and Hfr2 leading to F450 and *Flac*, respectively, may have been deleted from 0 to 3 kb, in spite of the evidence offered above that their parental F was the same as that studied by us. Type

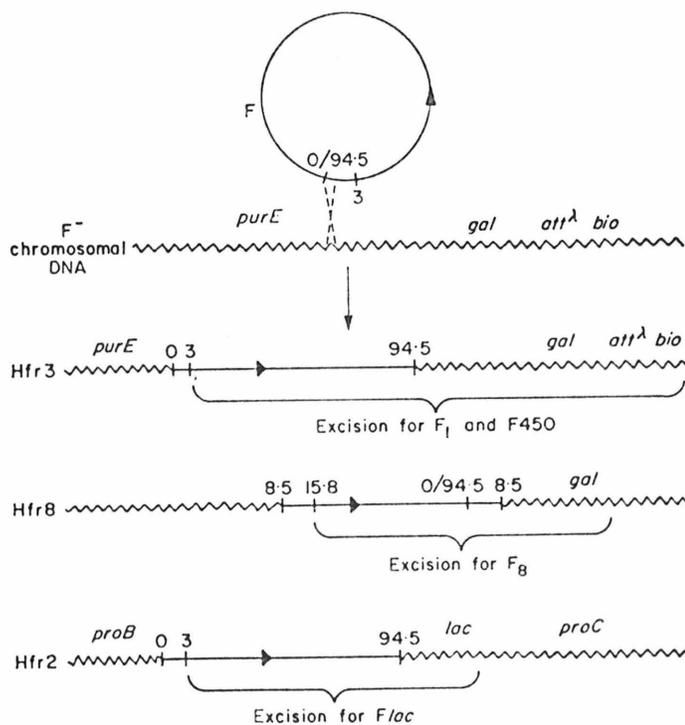


FIG. 15. Probable insertion and excision processes leading to the formation of Hfr's 2, 3 and the episomes F₁, F450, F₈, and *Flac*. The recombination process giving Hfr 3 is explicitly shown. The proposed structures of Hfr 8 and Hfr 2 are shown.

II excision would then lead to the structures observed. Nevertheless, the interpretation that insertion of the same sequence of F has occurred at different points in the bacterial chromosome seems inescapable.) The resolution of the electron microscope experiments is ± 500 nucleotides. To within this resolution, we may say that the point at 0/94.5 is a hot spot for recombination with the chromosome. This point is also implicated in the excision process leading to the formation of F Δ (0-14.5).

The insertion and excision points for the formation of the F₈ episomes are 8.5 and 15.8 respectively. Thus, more than one site on F can be used for insertions. It may be noted that a number of sites on the bacterial chromosome are available for insertion of F, but there is a preference for certain sites (Broda, 1967; Sanderson & Demerec, 1965; Matney, Goldschmidt, Erwin & Scroggs, 1964).

We wish to re-emphasize that the interpretations above are dependent on the following assumptions: (a) the F-primes arose by the classical process of insertion of F and excision as in Figure 1; (b) F as isolated from W1485 was the F that inserted to form the Hfr's from which the several F-primes arose; (c) the F-primes extracted have not changed structure during cultivation, storage and transfer.

Evidence which supports, but does not rigorously validate, these assumptions has been presented. A further structural study in which a set of Hfr's is derived from a single F⁺ male, and several F-primes are derived from each Hfr is clearly desirable.

This research has been supported by National Institutes of Health grant no. GM10991. One of us (P. A. S.) was the recipient of a National Institutes of Health Fellowship. Our indebtedness to A. D. Kaiser, D. Freifelder, A. Campbell, B. Low, D. R. Helinski, P. M. A. Broda, J. Gross, Y. Hirota, E. Lederberg and T. Fukasawa is indicated in the text. The initial stimulus for this work arose in the course of a discussion with J. S. Parkinson. This is contribution no. 4426 from the Department of Chemistry, California Institute of Technology.

REFERENCES

- Achtman, M., Willetts, N. & Clark, A. J. (1971). *J. Bact.* **106**, 529.
 Achtman, M., Willetts, N. & Clark, A. J. (1972). *J. Bact.* **110**, 831.
 Adhya, S., Cleary, P. & Campbell, A. (1968). *Proc. Nat. Acad. Sci., Wash.* **61**, 956.
 Bazarol, M. & Helinski, D. R. (1968). *J. Mol. Biol.* **36**, 185.
 Broda, P. (1967). *Genet. Res. Camb.* **9**, 35.
 Bruner, R. & Vinograd, J. (1965). *Biochim. biophys. Acta*, **108**, 18.
 Burton, A. & Sinsheimer, R. L. (1965). *J. Mol. Biol.* **14**, 327.
 Cairns, J. (1963). *Cold Spr. Harb. Symp. Quant. Biol.* **28**, 43.
 Campbell, A. M. (1969). *Episomes*. New York: Harper and Row.
 Clark, A. S. & Margulies, A. D. (1965). *Proc. Nat. Acad. Sci., Wash.* **53**, 451.
 Clayton, D. A. & Vinograd, J. (1967). *Nature*, **216**, 652.
 Cohen, S. N. & Miller, C. A. (1969). *Nature*, **224**, 1273.
 Crothers, D. & Zimm, B. (1965). *J. Mol. Biol.* **12**, 525.
 Davidson, N. & Szybalski, W. (1971). *The Bacteriophage Lambda*, ed. by A. D. Hershey, ch. 3. New York: Cold Spring Harbor Laboratory.
 Davis, R. W. & Davidson, N. (1968). *Proc. Nat. Acad. Sci., Wash.* **60**, 243.
 Davis, R. W. & Hyman, R. W. (1971). *J. Mol. Biol.* **62**, 287.
 Davis, B. D. & Mingioli, E. S. (1950). *J. Bact.* **60**, 17.
 Davis, R. W., Simon, N. & Davidson, N. (1971). *Methods in Enzymology*, vol. 21D, p. 413. L. Grossman, K. Moldave, eds. New York: Academic Press.
 Franklin, N. C. (1971). *The Bacteriophage Lambda*, ed. by A. D. Hershey, ch. 8. New York: Cold Spring Harbor Laboratory.
 Freifelder, D. (1968). *J. Mol. Biol.* **35**, 85.

- Freifelder, D., Folkmanis, A. & Kirschner, I. (1971). *J. Bact.* **105**, 722.
- Freifelder, D. & Meselson, M. (1970). *Proc. Nat. Acad. Sci., Wash.* **65**, 200.
- Gray, H. B., Jr., Bloomfield, V. A. & Hearst, J. E. (1967). *J. Chem. Phys.* **46**, 1493.
- Herbert, A. A. & Guest, J. R. (1968). *J. Gen. Microbiol.* **53**, 363.
- Hirota, Y. & Sneath, P. H. A. (1961). *Japan. J. Gen.* **36**, 307.
- Hradecna, A. & Szybalski, W. (1969). *Virology*, **38**, 473.
- Ihler, K., Achtman, M. & Willetts, N. (1972). *J. Bact.* **110**, 857.
- Jacob, F. & Wollman, E. L. (1961). *Sexuality and the Genetics of Bacteria*. New York and London: Academic Press.
- Kayajanian, G. & Campbell, A. (1966). *Virology*, **30**, 482.
- Laird, C. D., McConaughy, B. L. & McCarthy, B. J. (1969). *Nature*, **224**, 149.
- Lee, C. S., Davis, R. W. & Davidson, N. (1970). *J. Mol. Biol.* **48**, 1.
- Liedke-Kulke, M. & Kaiser, A. D. (1967). *Virology*, **32**, 465.
- Matney, T. S., Goldschmidt, E. P., Erwin, N. S. & Scroggs, R. A. (1964). *Biochem. Biophys. Res. Comm.* **17**, 278.
- McConaughy, B. L., Laird, C. D. & McCarthy, B. J. (1969). *Biochemistry*, **8**, 3289.
- Mizuuchi, K. & Fukasawa, T. (1969). *Virology*, **39**, 467.
- Ogawa, H., Shimada, K. & Tomizawa, J. (1968). *Molec. Gen. Genetics*, **101**, 227.
- Ohtsubo, E. (1970). *Genetics*, **64**, 189.
- Ohtsubo, E., Nishimura, Y. & Hirota, Y. (1970). *Genetics*, **64**, 173.
- Ray, D. S., Preuss, A. & Hofschneider, P. H. (1966). *J. Mol. Biol.* **21**, 485.
- Sanderson, K. E. & Demeree, M. (1965). *Genetics*, **51**, 897.
- Scaife, J. (1967). *Ann. Rev. Microbiol.* **21**, 601.
- Sharp, P. A., Cohen, S. N. & Davidson, N. (1973). *J. Mol. Biol.* in the press.
- Sharp, P. A., Hsu, M.-T. & Davidson, N. (1972). *J. Mol. Biol.* **71**, 499.
- Signer, E. R., Beckwith, J. R. & Brenner, S. (1965). *J. Mol. Biol.* **14**, 153.
- Taylor, A. L. (1970). *Bact. Rev.* **34**, 155.
- Taylor, A. L. & Trotter, C. D. (1967). *Bact. Rev.* **31**, 332.
- Vinograd, J., Lebowitz, J., Radloff, R., Watson, R. & Laipis, R. (1965). *Proc. Nat. Acad. Sci., Wash.* **53**, 1104.
- Wallace, B. J. & Pittard, J. (1967). *J. Bact.* **93**, 237.
- Wang, J. C. (1970). *Biopolymers*, **9**, 489.
- Westmoreland, B. C., Szybalski, W. & Ris, H. (1969). *Science*, **63**, 1343.
- Willetts, N. & Achtman, M. (1972). *J. Bact.* **110**, 843.
- Willetts, N. (1972). *Ann. Rev. Genetics*, **6**, in the press.

Chapter 2

Electron Microscope Heteroduplex
Studies of Sequence Relations
among Plasmids of E. coli. III.
Structures of F100, F152, F8 and
Their Derivatives.

This chapter is a preprint to be
submitted for publication in
Journal of Molecular Biology

Electron Microscope Heteroduplex Studies of Sequence Relations Among
Plasmids of *E. coli*. III. Structure of F100, F152, F8 and Their
Derivatives*

E. OHTSUBO and M. T. HSU

Department of Chemistry
California Institute of Technology
Pasadena, California USA 91109

Summary

The genetic and physical structures of F100, F152, F8 and some of their derivatives were analyzed. Deletions in the bacterial DNA sequence in some of them were identified and mapped. The physical positions of several of the genetic markers of *E. coli* between fep and uvrB were determined. A method was developed to reconstruct the original episomes from their deletion variants. The results confirm the history that F100 and F152 were derived from the same Hfr. A new episome, F80, containing genetic markers between fep and gal was reconstructed from F8. The formation of this new episome suggests that there is a hot spot in the *E. coli* chromosome for recombination with a particular F sequence to which we have assigned the coordinates 0.04 F.

1. Introduction

Several classical episomes carrying gal genes have been used extensively in genetic analysis. F100 (or F₁ gal) and F152 (or F₂ gal) were isolated from Hfr P3 by Jacob and Wollman (Bachmann, 1972). F8 was isolated from Hfr 8 by Hirota and Sneath (1961). Because of

*Contribution No. 4760

their usefulness, they have been widely distributed to many laboratories since their isolation. The discovery that spontaneous deletions occur in many F-prime factors (Low, 1972) raises the concern that the sequence of these episomes may have diverged from laboratory to laboratory.

Indeed, preliminary evidences reported previously (Sharp *et al.*, 1972) suggested that several descendants of F100 and F152 are probably deletion variants of the original episomes. In the present work, the genetic and physical structures of several descendants of F100, F152, and F8 were studied. Deletions in some of them were identified and mapped. The physical locations of several bacterial genes between fep and uvrB were determined. A method was developed to reconstruct the original episome from the deletion variants. This method may also be a potential tool for physical mapping of bacterial genes.

2. Material and Methods

(a) Bacterial strains

Bacterial strains used in this study are listed in Table 1.

(b) Media

Penassay broth (Difco) was usually used for growing bacterial cultures for genetic experiments. Bacteria were grown in tryptone broth containing 1 μ g/ml thiamine for extraction of episomal DNA. L plates (Lennox) were used as a complete solid medium. Davis minimum agar plates (Davis and Mingioli, 1950) were used for genetic analysis. They were supplemented with appropriate amino acids, vitamins, and 0.2% glucose (or 1% galactose), depending on the genetic

marker to be tested. EMB plates were used for gal and mal markers.

(c) Transfer of F'gal's

In experiments involving the transfer F'gal's into other recipients we avoid the use of EMB galactose plates as selective media. This is because bacteria carrying smaller F'gal's resulting from spontaneous deletion are usually more recognizable as gal⁺ colonies on EMB galactose plates and therefore may be preferentially picked up for analysis.

Consequently, the following procedure was used instead. All of the transfers of F'gal's were done using N23-53 (F⁻ gal⁻ trp⁻ arg⁻ str^r recA⁻) either as a donor or as a recipient. This bacterium has an extremely slow growth rate and therefore can be easily selected for as a recipient or selected against as a donor. Donor and recipient bacteria were grown in penassay broth at 37°C overnight. They were mixed in a ratio of about 5 to 1 and incubated for another two hours at 37°C. The mating mixture was then plated with appropriate dilution on an L plate which allows every bacterium plated to grow and form a colony. Tiny colonies were picked up for examination when N23-53 was used as recipients while large colonies were selected when N23-53 was used as donor. Over 80% of the bacteria selected by this method had received the F'gal's. The condition of mating used in reconstruction experiments is described in the text.

(d) Testing of bacterial markers on F'gal's

N23-53 strains carrying various F-prime factors were usually used as donors in testing the transferability of episomal markers. Since N23-53 is recA⁻, the contribution due to chromosome mobilization can be neglected.

The cross-streaking method (Lederberg, 1949) was generally used. Donor strains carrying various F'gal's were cross-streaked against suitable recipients on the appropriate selective plates described in the reference for each marker. The markers tested using this method are shown in Table 2. Genetic analysis of supE is more complicated and is described in the text.

The procedure for other markers are given below:

(i) LeuS and SerS (Low et al., 1971):

The recipients used are KL231 and KL229 respectively. They are temperature sensitive mutants in these two genes. The recipients were grown at 30°C, cross-streaked with N23-53 donors on Davis minimum plates and then incubated at 42°C. Growth of recipients in the cross-streaked area at the restrictive temperature indicates the presence of these markers in the F-prime factor tested.

(ii) tolAB (Nomura and Witten, 1967): -

The F'gal to be tested was transferred into N0712 and N0570 and the F-ductants were tested for sensitivity to colicin E2 spontaneously excreted from N052.

(iii) pgl (Kupor and Frankel, 1969):

F'gal in N23-53 were transferred into SA291 and the gal⁺ F-ductants were examined for "blue" character on EMB maltose plate.

(iv) uvrB (Ogawa et al., 1968):

The F-ductants in (iii) were examined for their uv sensitivity according to the method used by Ogawa et al.

-5-

(v) att^λ (Rothman, 1965):

The same F-ductants used in the previous test were cross-streaked against a phage λ lysate on EMB maltose plate. A clear inhibitory zone on the cross-streaked region indicates the absence of att^λ in F'gal's.

(vi) lip (Herbert and Guest, 1968)

A direct test of lip⁺ transferability of reconstructed episomes harbored in PL225 and KRO is performed as follows: To select against the donors, a transfer defective R factor, R100-31, which cannot be complemented by the F transfer system and carries str, chl, sul and drug resistance markers was introduced into the recipient, AB1325 lip 13/P1 transduction. The donors were replica plated onto lip selective plates seeded with AB1325 (R100-31) and containing 25 μg/ml of chloramphenicol.

(e) Isolation of episomal DNA

Due to the large size of ^{the}episomes studied, the following modifications of the previous procedures (Sharp et al., 1972) were used. The lysozyme treated cells were lysed by adding SDS to a final concentration of 0.2% and incubated at 37°C for several minutes until the solution became clear. The DNA was then sheared rather gently to avoid breakage of large DNA by two passages through a syringe as described in Sharp et al. (1972). The DNA solution was always filtered through glass wool before the CsCl-ethidium bromide centrifugation. For small episomes of size between one and two times that of F sizes, the volume of bacterial culture used was usually reduced to 500 ml so that three different episomes could be isolated at the same time.

(f) Electron microscope analysis of heteroduplexes

The technique for spreading heteroduplex molecules has been described in Sharp *et al.* (1972). The length of single-strand DNA was measured against the ϕ X viral DNA internal standard. The length of homoduplex DNA in the same grid as the heteroduplex was used to fix the size of the duplex region in heteroduplex molecules.

(g) Methodology for identifying F and bacterial sequences in an episome

The standard procedure used is to analyze the structure of heteroduplexes with F and with F8-33. A heteroduplex with F is useful in measuring the size of bacterial DNA carried in an episome. F8-33 carries two characteristic small insertions at coordinates 91.0 and 35.2* on the physical map of F (Sharp *et al.*, 1972) as shown in Fig. 1. These two insertions can serve as physical markers to map the F or the bacterial sequence in the F-prime factors.

(h) Nomenclature of F-prime factors

The nomenclature of F-prime factors studied in this paper follows that suggested by Low (1972). F100 is equivalent to F_{1gal} and F152 is the same as F_{2gal}. F_{1s} and F450 (Sharp *et al.*, 1972) are renamed as F100-1 and F100-2 because they are shown to be deletion derivatives of F100 (see text). F₈(1), F₈(2) and F₈(N33) studied in Sharp *et al.* (1972) are renamed as F8, F8-2 and F8-33 respectively.

3. Results and Discussions

(a) Molecular weights of F'gal's

The molecular weights, calculated from measured duplex lengths, of the episomes studied are displayed in Table 4. About 12 molecules were

measured for each episome. F DNA was used as an internal standard. All the DNA preparations contain various kinds of small circular DNA species. The fraction of these unknown plasmid DNA's are about 10% or less by number of the total DNA isolated.

(b) Analysis of bacterial markers carried by F'gal's

The F-prime factors were tested for the ability to transfer various genetic markers in the region between purE and uvrB of the E. coli some (see Material and Methods). The results are summarized in Table 4.

(c) Physical structures of F'gal's

The sequence relations between the several episomes studied are rather complex. It will be conducive to clarity, we believe, to first present the final results and then to give the evidences that leads to these interpretations.

Figure 2 is a summary of the genetic and physical structures of the various episomes studied. The structures of the bacterial DNAs are shown in Fig. 2a. Sequences deleted are indicated by a dotted line. For example, F100-1 and F100-2 are deletion mutants of the "original" F100. The bacterial sequence between coordinates 6.9 kb and 115.8 kb is deleted in F100-1. This corresponds to the absence of lip, leuS, ubiF, fep markers in this episome. In F100-2, the sequence deleted is between 29.0 kb and 195.4 kb. The bacterial markers missing are between fep and sucAB.

The coordinate system we have chosen to use for the bacterial DNA of F100 and F152 and their derivatives is shown in the top line of Fig. 2a. The coordinate system for the bacterial DNA of F80 is the same as that for F100 and F152. For F8 and its derivatives the coordin-

ate system we use is different from that of F100. It is shown in the bottom line in Fig. 2a. We have taken the counterclockwise junction of bacterial DNA with F DNA in an episome as the origin of coordinates for the bacterial DNA.

The coordinates of the bacterial and the F sequences in an episome are identified by the letters B and F, respectively, following the coordinate number. For example, the deletions of F100-1 are between 6.9 B and 115.8 B in the bacterial DNA and between 0.0 F and 2.8 F in the F DNA.

The arrangements of F sequences in the episomes studied are shown in Fig. 2b. F100, F152, and their derivatives are all missing the F sequences with coordinates from 0.0 F to 2.8* F; the junctions with bacterial DNA occur at these two points. Similarly, in the F8 episomes, the F sequence, 8.5 F to 16.3* F, has been substituted by bacterial DNA. (The small insertion loops, ϵ and ζ at coordinates 35.2* F and 91.0 F that occur in some F8 derivatives are not shown in Fig. 2.) In F80, the F sequence substituted is between 0.0 F and 16.3* F.

The evidence leading to the structures shown in Fig. 2 is described below:

(A) F152 (F₂gal)

F152 was originally isolated by Wollman (Bachmann, 1972). Its genetic properties have been intensively studied by Herbert and Guest (1968). It transfers gal as a proximal marker and lip as a distal marker.

Previous efforts to isolate this episome from W3101(F152) and PA106 (F152) resulted in finding only a deleted F (Sharp *et al.*, 1972). Two episomes were isolated from a new source, KLF253(F152), kindly supplied by B. Low. Their physical structures are described below:

* Revised coordinates of those in Sharp *et al.* (1972). R. C. Deonier, Private communication.

(i) F152

A large episome of about 310 kb in length was isolated from a stab of KLF253(F152). Genetic analysis of this episome (see Table 4) showed that it carries all the bacterial markers tested between fep and gal. We believe this is the original F152 episome. Evidence supporting this assumption will be presented later (see sec. D).

F152 DNA isolated from KLF253(F152) carries a segment of bacterial DNA of length 218.4 kb, with assigned coordinates from 0.0 B to 218.4 B. It is deleted in the F sequence between 0.0 F and 2.8 F (see Fig. 2a and b). These results were obtained by analyzing the structure of F152/F and F152/F8-33 heteroduplexes as shown in Fig. 3a and b respectively. (An electron micrograph of 3a is shown in Plate L) The two characteristic insertion loops c and $\zeta\epsilon$, of F8-33 were used to identify the F sequence deleted (see Material and Methods).

The bacterial DNA of F8-33 is homologous to a part of the bacterial DNA of F152, with coordinates from 195.4 B to 217.2 B (see Fig. 2a). This is the double-strand DNA segment be in Fig. 3b. The bacterial DNA of F152 with coordinate between 0.0 B and 195.4 B is absent in F8-33 and is seen as the single-strand segment ab in Fig. 3b. Evidence that the sequence, 217.2 B - 218.4 B, is not present in F8-33 will be given in the next section. The shorter arm of the substitution loop in Fig. 3b, afb, is the F sequence in F8-33 with coordinate between 0.0 F and 8.5 F as mapped from the position of the insertion loop c. The insertion

or deletion loop (I-D loop) at point e contains F sequences in F152 between 2.8 F and 16.3 F. The sequence, 2.8 F - 8.5 F, present in these two single-strand segments can sometimes form base pairs as indicated by the dotted line. This pairing was seen only in incomplete structures, that is, in which one or both of the strands was an incomplete fragment.

(ii) F152-1, a deleted F152

The first stab of KLF253(F152) we received from B. Low was found to contain mostly gal⁻ bacteria when streaked on EMB gal plate. One of the few gal⁺ colonies was picked up and its genetic properties examined. The presence of an F-prime factor capable of transferring gal marker was confirmed. However, several bacterial markers which are supposed to be present in F152 were missing (see Table 4). The episome isolated has a length of 120 kb, considerably shorter than F152. We name it as F152-1.

The physical structure of this episome is summarized in Fig. 2. It carries a deletion of ^{an} F sequence with coordinate from 0.0 F to 2.8 F (see Fig. 2b) as shown by the structure of heteroduplexes F152-1/F and F152-1/F8-33 (see Material and Methods). The bacterial sequence present in this episome is 26.5 ± 1.3 kb long. It is derived from F152 by a single large deletion in the chromosomal DNA from 17.9 B to 207.2 B as shown below. When F152-1 was hybridized with F152, a single deletion loop of 190.0 ± 4.4 kb long was observed in the heteroduplex (see Fig. 4 and Plate II). The position of the deletion is mapped by analyzing the structure of the F152-1/F8-33 heteroduplex as shown in Fig. 5a, b. In Fig.

5a, the duplex region be with a length of 10.0 kb long is the homologous bacterial DNA of F152-1 and F8-33. (An electromicrograph of Fig. 5a is shown in Plate III.) The length of this segment is about 11 kb shorter than that in F152/F8-33 heteroduplex (see Fig. 3b). It has coordinates from 16.0 B to 26.6 B in F8-33. This corresponds to the bacterial DNA of F152 with coordinate from 207.2 B to 217.2 B (see Fig. 2a). Point b in F152-1 with coordinate 207.2 B therefore is one of the deletion end points in F152-1. Since F152-1/F152 shows a single deletion of 190 kb long, the other deletion end point can be readily calculated to be $207.2 - 190 = 17.2$ B. This can also be deduced from the length of the single-strand segment ab. As can be seen from Fig. 2a this is the bacterial DNA of F152-1 with coordinates from 0.0 B to 17.9 B. The latter data are more accurate and are therefore preferred. The bacterial sequence missing in F152-1 is thus mapped between 17.9 B and 207.2 B. This segment therefore contains the genetic markers between fep and nadA, which are absent in F152-1. NadA has been mapped to lie between 14.2 B and 15.9 B in F8 episomes (Sharp et al., 1972). This corresponds to the DNA of F152 with coordinate between 204.8 B and 206.5 B which is deleted in F152-1. This is in good agreement with the genetic analysis that nadA but not aroC is missing in F152-1 (see Fig. 2a).

The structure shown in Fig. 5b is sometimes observed for F152-1/F8-33 heteroduplex. It is formed from structure 5a by the base pairing of the F sequence, 2.8 F - 8.5 F, in the loop

leaving the duplex at point e and the single-strand segment af'g'b, as indicated by the dotted line in Fig. 5a. From this structure, the small bacterial DNA segment of F152-1 (ef in Fig. 5b) with coordinate from 217.2 B to 218.4 B that extends clockwise beyond the bacterial DNA present in F8 can be identified. Presumably this sequence is also present in F152.

(B) F100

F100 and F152 were said to be independently derived from Hfr P3 (Bachmann, 1972). F100 differs from F152 by carrying bacterial markers att^λ, bio and uvrB (Low, 1972) in addition to those present in F152. The structures of two episomes F100-1 (F_{1g}) and F100-2 (F450) have been studied in our previous work (Sharp *et al.*, 1972). Preliminary results indicated that they were probably different deletion variants of F100. As will be described below, the deletions in these two episomes are confirmed and mapped.

(i) F100-1

As described briefly in the previous paper, this episome carries a deletion in F sequence from 0 F to 2.8 F. The bacterial DNA is 136.7 ± 3.5 kb in length, considerably shorter than that of F152. This is due to a large deletion in the bacterial DNA of F100-1 between 6.9 B and 115.8 B (see Fig. 2a). The structures of the F100-1/F152-1 and the F100-1/F152 heteroduplexes that lead to this interpretation and shown in Figs. 6 and 7 respectively. (An electromicrograph of Fig. 6 is shown in Plate IV.) In Fig. 6 a substitution and an I-D loop can be seen. The I-D loop at e is the bacterial DNA of F100-1 from 218.4 B to 249.0 B. It contains sequences of att^λ, bio, and uvrB markers which are missing in F152-1. The double-strand segment be is the common bacterial

sequence of the two episomes with coordinates between 207.2 B and 218.4 B. The smaller arm of the substitution loop is the bacterial DNA of F152-1 from 6.9 B to 17.9 B missing in F100-1. The larger arm of this loop is the bacterial DNA of F100-1 from 115.8 B to 207.2 B that is missing in F152-1. The bacterial sequence, 0 B-6.9 B, is present in both episomes. Consequently, the bacterial DNA deleted in F100-1 is mapped between 6.9 B and 115.8 B. This result is confirmed from the structure of F100-1/F152 shows in Fig. 7 in which the sequence deleted in F100-1 is seen as a single-strand loop at point a, about 102 kb from the point b.

An important result of the structure of F100-1 is that DNA sequence around the integration point (0B/0F in Fig. 7, see also Fig. 17) is the same as that in F152. This is consistent with the history that F100 and F152 were derived from the same Hfr.

(ii) F100-2 (F450)

The structure of this episome has been studied extensively by Sharp et al. (1972). It carries a deletion in F sequences from 0 F to 2.8 F. The bacterial DNA is 80.0 kb in length. The evidence to be described below suggests that it is probably derived from the "original" F100 by a single deletion in the bacterial sequence from 29.0 B to 195.4 B (see Fig. 1a). A heteroduplex of this episome with F100-1 is depicted in Fig. 8 and Plate V. Point a in F100-1 DNA is readily identified as the deletion end point of F100-1 and has the coordinate 6.9 B/115.8 B. From this the coordinate at point b is calculated to be at 29.0 B/195.4 B from

the lengths of the two single-strand arms of the substitution loop. The bacterial DNA deleted in F100-2 is therefore from 29.0 B to 195.4 B. This is confirmed by the structure of the F100-2/F152-1 heteroduplex as shown in Fig. 9 and Plate VI.

(C) F8 and F8-3

As shown in the F152/F8-3 heteroduplex, the bacterial sequence of F8-33 is homologous to that of F152 from 217.2 B to 195.4 B. F8 carries a piece of bacterial DNA of length 5.9 kb to the left of the bacterial DNA of F8-33 (see Fig. 2a). That this DNA is the normal bacterial sequence to the left of sequence 5.9 B - 26.6 B is shown by the analysis of the F100-1/F8 and F100-1/F8-33 heteroduplexes (see Figs. 10a and 10b; a micrograph of Fig. 10a is shown in Plate VII). The homologous

bacterial sequence, be, in the F100-1/F8 heteroduplex is about 5.9 kb longer than the corresponding sequence in the F100-1/F8-33 heteroduplex. This indicates that the bacterial DNA of F8 is a continuous piece of normal E. coli chromosomal DNA.

F8-3 is a derivative of F8 obtained from Fukasawa. Genetic analysis showed that it does not carry ^{the} tolAB, nadA, and aroG markers which were shown to be present in F8 (Sharp *et al.*, 1972). Analysis of the structure of the F8-33/F8-3 heteroduplex as shown in Fig. 11 showed that it is derived from the original F8 by a single deletion of bacterial DNA between 0 B and 20.6 B in the F8 coordinate system. Since the physical positions of the aroG and nadA genes have been mapped between 0 B and 18.6 B (Sharp *et al.*, 1972), this result is consistent with the genetic properties of this new episome.

D] Reconstruction of F152 and F100

The interpretation of the electron microscope studies in the previous section depends on the assumption that the F152 episome isolated from KLF 253 (F152) (hereafter abbreviated as F152K) carries the normal bacterial sequence of E. coli between fep and gal and is the same as that of the original F152. To test this assumption, we compare F152K with F152 and F100 that have been reconstructed from their deletion derivatives by rescuing the sequences deleted by recombination with wild-type bacterial chromosomes. The results to be presented below support the assumption that F152K is the same as the original F152.

(a) Isolation of reconstructed F152 from F152-1 and comparison of it with F152K

The process for reconstruction of F152 from F152-1 is illustrated in Fig. 12. F152-1 which carries a large deletion in bacterial DNA was transferred into W3350 which carries double point mutations in gal but is otherwise normal. Recombination between F152-1 and the W3350 chromosome in regions 1 and 2 or in regions 1 and 3 will result in formation of an episome carrying the sequence deleted in F152-1. These new episomes can be selectively transferred into PL225 by mating W3350 (F152-1) (gal⁻ 1, 2, recA⁺, str^S) with PL225 (recA⁻ str^R Δ (nadA⁻ gal)) and selecting for nadA⁺ str^R recombinants. The frequency of such recombination is about 5×10^{-5} per donor cell. Among 135 nadA⁺ str^R colonies selected, 75% of the recombinants are gal⁻ resulting from the recombination in regions 1 and 3. All the recombinants are capable of transferring the lip⁺ marker, which is deleted in F152-1, with high efficiency. For convenience in the following discussion, we will call these new episomes F152R.

The episomes of each recombinant was transferred into N23-53 (gal⁻ recA⁻) in order to test the bacterial markers carried. Some of the results are given in Table 4. F152R is indistinguishable from F152K for all the markers tested. An episome was extracted from one of the N23-53 (F152R) strains and hybridized with F152K. No heteroduplex of any sort could be observed in the electron microscope. The structure of the F152R/F100-1 heteroduplex is essentially the same as that of the F152K/F100-1 heteroduplex. Therefore F152R and F152K are the same by the heteroduplex criterion. We suggest that the chance for having the same deletion or insertion in F152 giving rise to F152K and in the W3350 chromosome giving rise to F152R is rare. Consequently F152K is the same as the original F152 episome.

The recombination process leading to the reconstruction of F152 from F152-1 is more dramatically demonstrated by reconstructing a new F152 episome carrying a deletion originally located in the bacterial chromosome (see Fig. 12). F152-1 was transferred into FRAG-5 which carries a deletion in the kdp gene. The reconstructed episome was then selectively transferred into PL225 by selecting nadA⁺ str^r recombinants. As in the case of the F152R episome, both gal⁺ and gal⁻ recombinants were obtained. The frequency of gal⁺ bacteria was 43/52. All the recombinants can transfer lip with high efficiency. Genetic analysis of these new episomes in N23-53 showed that the kdp marker is missing. The physical structure of one of the new episomes, which we name as F152-2, was analyzed. The sole nonhomology feature observed in the F152-2/F152K heteroduplex is an I-D loop 3.9 ± 0.3 kb long. From the heteroduplex F152-2/F100-1 (see Fig. 7 and Plate VIII) the position of this I-D loop was mapped to be between coordinates 151.7B and 155.7B, assuming it is a deletion. We believe this is the kdp deletion which is originally present in FRAG-5 chromosome.

The simple structure observed in the F152K/F152-2 heteroduplex again confirms the previous conclusion that the bacterial DNA carried in F152K is the normal bacterial sequence.

(b) Reconstruction of F100 from F100-2

An independent check of the previous results is to compare F152K with a reconstructed F100. The process for reconstruction of F100 is closely related to the genetic analysis of supE marker in the episomes studied. Since analysis of the supE marker in an episome is complicated,

we will first discuss how this is done. In the following discussion, the F100-2 episome will be used as an example.

When F100-2 was transferred into the bacterium KRO which carries an amber mutation in the trp gene, no trp⁺ progeny were obtained. Therefore the phenotype of the F100-2 episome is supE⁻. Since the supE⁻ allele is the normal genotype, it is difficult to determine whether the supE⁻ phenotype of F100-2 is due to the presence of a normal supE⁻ allele or ^{to} the deletion of the supE sequence. To differentiate these two possibilities, a marker rescue experiment similar to those discussed in the previous section is used. F100-2 was transferred into χ 7026 which carries the supE⁺ allele. Two possible structures of the χ 7026 (F100-2) heterogenotes are shown in Fig. 13a and b depending on whether the supE⁻ allele is present in F100-2. As shown in Fig. 13a., if the supE⁻ allele is not present in F100-2, a supE⁺ episome resulted from the recombination between F100-2 and χ 7026 chromosome will be always lip⁺. On the other hand, if F100-2 carries the supE⁻ allele, as shown in Fig. 13b, some of the supE⁺ episomes will still carry the lip deletion due to the two cross-overs in regions 2 and 3. When χ 7026 (F100-2) was mated with KRO (str^r trp amber, recA⁻), 84 trp⁺ str^r recombinants isolated were all found to be capable of transferring lip⁺ at high frequency. To this extent, we conclude that supE allele is not present in F100-2. However, it is possible that F100-2 does carry the supE allele but the sequence in region 2 in Fig. 13b is so short that the frequency of a supE⁺ lip⁻ _{as} episome is so low/to escape our limited survey. We think this is unlikely judging from the following facts: (i) sucAB and kdp markers which are known to map between supE and gal (Taylor and Trotter, 1972) are absent in F100-2 (see Table 4). (ii) F100-2 contains bacterial sequences

with coordinates between 0.0 B and 29.0 B and between 195.4 B and 249.0 B (see Fig. 2a). The latter segment is homologous to F8 which does not carry the supE allele (Signer *et al.*, 1965).

The results of the supE marker test for other F'gal's are given in Table 5. Thus with the limited survey shown in Table 5, F152-1 does not seem to carry the supE sequence while F100-1 and F153 do. The result for F8 is more complicated. A detailed discussion of this topic will be given in the next section.

As shown in Fig. 13a, the F100 episome can be reconstructed from F100-2 by rescuing the supE⁺ marker. Therefore the episome present in str^r trp⁺ recombinants from the cross between χ 7026 (F100-2) and KRO must be the desired reconstructed F100 episome. The reconstructed F100 (hereafter abbreviated as F100R) was transferred into N23-53 for genetic analysis. All the markers between fep and uvrB tested are present in this episome. The episomal DNA isolated is almost twice as large as F100-2. The structure of the F100R/F152-2 heteroduplex is shown in Fig. 14 and Plate IX. Three I-D loops were observed: the I-D loop at point a with a length of 30 kb is identified as the sequence containing att ^{λ} , bio, and uvrB which is missing in the F152 episome; the I-D loop at point b with a length of 4 kb is the kdp deletion of F152-2; the I-D loop at point c mapped at coordinate 211.7B is an unexpected structure. Its length is similar to that of the IS1 insertion found in some strong polar mutations of bacterial genes (Hirsch *et al.*, 1972). Disregarding the small I-D loop at point c, the structure of F100R is that expected for the hypothetical F100 episome. It carries att ^{λ} , bio, and uvrB sequences in addition to those present in

F152. The sequence near the counterclockwise F-chromosomal DNA junction in F100R is homologous with F152 indicating that they were derived from the same Hfr. This result also reaffirms our previous conclusion that F152K carries the normal bacterial sequence and is the same as the original F152.

Since the reconstructed F100 episome carries a small insertion or deletion of unknown nature, we shall call it F100-3 to distinguish it from the "original" F100.

E) F80, a new episome carrying supE⁺, derived from F8.

F8 carries bacterial DNA from 188.8B to 217.2B on the physical map of F152 (see Fig. 2a). Therefore the supE sequence is not present in F8. However, in the cross between χ 7026 (F8) and KRO (see sec D and Table 5), trp⁺ str^r colonies were found at a low frequency. These are presumably due to the suppression of trp amber mutation by the presence of supE⁺ in the episome transferred into KRO. These bacteria can transfer the lip⁺ marker at high frequency, suggesting that the lip⁺ marker is also present in the episome in the recombinants. As will be discussed below, this is due to the formation from F8, of a new episome, carrying supE⁺ and lip⁺ markers, which we name as F80. The proposed mechanism for the formation of F80 is illustrated in Fig. 15. The transformation of F8 into F80 is achieved in two steps: (i) Integration of F8 into the bacterial chromosome of χ 7026 through the recombination between homologous bacterial sequences of F8 and the χ 7026 chromosome. (ii) Subsequent excision of the integrated F8 episome through recombination the F sequence of F8 at 0.0 F and a hot spot for F integration between purE

and lip in the $\chi 7026$ chromosome. We postulate that this hot spot is the same site used for the integration of F at 0.0 F to form Hfr P3.

From this model, the new episome F80 is predicted to have the following properties: (i) It should carry bacterial markers between fep and gal. (ii) It should be deleted in the F sequences between 0.0 F and 16.3 F. (iii) It should carry the bacterial sequences of F152 from 0.0 B to 217.2 B.

These properties are indeed observed as shown in Table 4 and by the structures of several heteroduplexes. The F80/F heteroduplex shows that F80 carries a deletion of 16.3 kb of F sequences (see Fig. 16a). The location of the deletion is mapped to be between 0.0 F and 16.3 F from the positions of the insertion loops c and ζ in the F80/F8-33 heteroduplex (Fig. 16b and Plate X). The bacterial sequence of F80 is homologous to that of F152 as shown by the F80/F152-2 heteroduplex in which the kdp deletion can be readily recognized. This conclusion is also confirmed by analysis of the F80/F100-1 heteroduplex (Fig. 16c and Plate XI).

4. Further Discussions

The results presented show that deletions have occurred in several of the frequently used F-prime factors such as F100, F152, and F8. The occurrence of these deletions is useful for localizing the physical positions of several bacterial markers between fep and uvrB. They are shown in the top line of Fig. 2a. In general these positions are in good agreement with the E. coli linkage map (Taylor and Trotter, 1972). The order of some markers such as fep, leuS and lip was not studied by us. They are shown in the order given in the E. coli linkage map.

The physical size of DNA between purE and uvrB is at least 230 kb long, enough to code for 230 ordinary sized genes. Only about 50 genes

have been mapped in this region. There probably are many others.

The physical length corresponding to the genetic unit, minute, can be calculated approximately from the relative physical positions of bioB, att^λ, galK, aroG, nadA, and kdpC to be about 46 kb. From this the molecular weight of the E. coli chromosomes can be calculated to be about 2.7×10^9 daltons. This agrees well with the molecular weight, 2.5×10^9 daltons, determined by autoradiographic measurement (Cairns, 1963, E. coli M.W. = $22 \times T_2$. M.W. of unglucosylated T_2 is 1.1×10^8 , Kim and Davidson, 1973).

The reconstruction of F152-2 carrying kdp deletion illustrates a general and potentially very useful method for physical mapping of bacterial genes. This method can be extended to the mapping of genes inactivated by Mu phage insertion (Hsu and Davidson, 1972).

The episomes F100 and F152 and their derivatives were found to have the same integration site not only on the F DNA but also on the bacterial chromosome. This is consistent with the history that they were derived from the same Hfr. The hypothetical process for the formation of F100 and F152 is illustrated in Fig. 17. It is interesting to note that although the excision point (2.8 F) on F DNA is the same in both cases, the excision point on the bacterial chromosome is different for the two episomes.

The isolation of the new episome, F80, from F8 (see Fig. 15) indicates that there is a hot spot between purE and lip in the E. coli chromosome responsible for the integration of F at coordinate 0.0 F to form Hfr P3. We speculate that this site is also responsible for the formation of Hfr C

and a transposed Hfr, Hfr EC29 (Beckwith et al., 1966) which have the origins located between purE and lip and have the same polarity as HfrP3. Rigorous proof of this hypothesis can be obtained by analyzing the physical structures of F-prime factors derived from Hfr C and Hfr EC29, and the structures of F-prime factor carrying DNA sequence between purE and lip.

Acknowledgement

We wish to express our gratitude to Professor Norman Davidson for his generosity in providing his laboratory space and funds (United States Public Health Service, Grant GM 10991). We would also like to thank him for reading the manuscript in preparation. We thank Drs. P. Broda, A. Campbell, W. Epstein, D. Freifelder, T. Fukasawa, J. R. Guest, B. Low, M. Nomura, E. Signer, and I. Young for providing the bacterial strains used in this work.

Table 1

<u>Strain</u>	<u>Plasmid</u>	<u>Chromosomal genotype</u>	<u>Source</u>
B583	F100-2	$\Delta(\text{gal-att}^{\lambda}\text{-chlA})$	D. Freifelder
MR44	F100-1	<u>galE galT recA spc</u>	E. R. Signer
JE5606	F153	<u>gal lac met arg his</u> <u>leu recA str</u>	E. Ohtsubo
KLF253	F152	<u>gal recA pyrD try his tyr</u> <u>thi mal xyl mtl</u>	B. Low
ND6	F152-1	<u>gal recA pyrD try his tyr</u> <u>thi mal xyl mtl</u>	This paper
JE5303	F8	<u>recA str gal6 trp arg</u>	E. Ohtsubo
JE3100	F8-2	<u>thr leu gal2 lac52 pil</u> <u>fla str</u>	E. Ohtsubo
W4520	F8-3	<u>met</u>	T. Fukasawa
JE3513	F8-33	<u>thr leu gal2 lac52 pil</u> <u>fla str</u>	E. Ohtsubo
W3747	F13	<u>met</u>	E. Ohtsubo
N23-53	F ⁻	<u>recA str gal6 trp arg</u>	E. Ohtsubo
PL225	F ⁻	<u>recA str thi $\Delta(\text{gal-nadA})$</u>	T. Fukasawa
W3350	F ⁻	<u>gal 1,2</u>	E. Ohtsubo
PB314	F ⁻	<u>lac.purE thi str tsx</u>	P. Broda
KL231	F ⁻	<u>leuS(ts) thyA str</u>	B. Low
AB1325 <u>lip13</u>	F ⁻	<u>tsx proA lacY gal purB</u> <u>his str mtl xyl thi lip13</u>	J. R. Guest
FRAG-5	F ⁻	<u>lacZ\times82 (amber) gal rha</u> <u>thi kdpABC-5</u>	W. Epstein

Table 1 (continued)

<u>Strain</u>	<u>Plasmid</u>	<u>Chromosomal genotype</u>	<u>Source</u>
AN102	F ⁻	<u>leu proC trp thi fep</u>	I. Young
AN146	F ⁻	<u>ubi-411 gal ura</u>	I. Young
W620	F ⁻	<u>gltA galK suc ura str</u>	J. R. Guest
FRAGG-1	F ⁻	<u>lacZx82 gal rha thi</u> <u>gltA</u>	W. Epstein
WGA <u>suc23</u>	F ⁻	<u>gal trp suc23 (sucA)</u>	J. R. Guest
W3110 <u>suc17</u>	F ⁻	<u>suc17 (sucB)</u>	J. R. Guest
NO712	F ⁻	<u>gal tolIII thr leu</u> <u>leu proA his arg thi</u> <u>lac xyl ara mtl str</u>	M. Nomura
NO570	F ⁻	<u>gal tolII thr leu</u> <u>lac str</u>	M. Nomura
W3102	F ⁻	<u>galK</u>	E. Ohtsubo
W3995	F ⁻	<u>galE str</u>	E. Ohtsubo
SA291	F ⁻	<u>his str Δ(gal-chlA)</u>	A. Campbell
KL229	F ⁻	<u>serS thyA35 str</u>	B. Low
KRO	F ⁻	<u>lac4680 (deletion) trp8</u> <u>(amber) str recA</u>	E. Signer
x7026	F ⁻	<u>Δ (lac-pro) supE⁺</u>	W. Epstein
NO52	ColE	<u>str</u>	M. Nomura
ND7	F152-2	<u>recA str thi Δ(gal-nadA)</u>	This paper
ND8	F100-3	<u>lac4680 (deletion)trp8</u> <u>(amber) str recA</u>	This paper
ND9	F80	<u>lac4680 (deletion) trp8</u> <u>(amber) recA</u>	This paper

Table 2

<u>Genetic Marker</u>	<u>Recipient</u>	<u>Reference</u>
<u>purE</u>	PB314	Stouthamer <u>et al.</u> , 1965
<u>fep</u>	AN102	Cox <u>et al.</u> , 1970
<u>lip</u>	AB1325 <u>lip13</u>	Herbert and Guest, 1968
<u>ubiF</u>	AN146	Young <u>et al.</u> , 1971
<u>kdp</u>	FRAG-5	Epstein and Davies, 1970
<u>sucA</u>	WGA <u>suc23</u>	Herbert and Guest, 1968
<u>sucB</u>	W3110, <u>suc17</u>	Herbert and Guest, 1968
<u>nadA</u>	PL225	Taylor and Trotter, 1967
<u>aroG</u>	PL225	Wallace and Pittard, 1967
<u>gal</u>	W3102, W3995	Adler and Kaiser, 1963
<u>bio</u>	SA291	Rolfe and Eisenberg, 1968

Table 3

	purE	fep	leuS	lip	supE ^a	ubiF	kdp	gltA	suc _B ^A	tol _B ^A	nadA	aroG	gal	pgl	att ^λ	bio	uvrB	SerS
F152	-	+	+	+	±	+	+	+	+	+	+	+	+	-	-	-	-	
F152-1	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	
F152-2			+	+			-	+	+				± ^b	-	-	-	-	
F100-1	-	-	-	-	±	+	+	+	+	+	+	+	+	+	+	+	+	-
F100-2	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-
F100-3				+	+			+	+				+					
F80	-			+	+								+					
F8	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-
F8-2				-					-	+	+	+	+					
F8-3				-					-	-	-	-	+					
F153	-			-			+	+	+		+	+	+	+	+	+	+	
F13	+			-						-	-	-	-	-	-	-	-	

a) The sign "±" means that the sequence is present in the episome but the phenotype is supE⁻.

b) Some of the episomes analyzed are gal⁻. See Section D.

Table 4

DNA	Mol wt $\times 10^{-6}$ daltons	kb
F152	205.1 \pm 5.0	310 \pm 7.6
F152-1	79.6 \pm 1.3	120.2 \pm 1.9
F100-1	151.9 \pm 2.3	229.5 \pm 3.4
F100-2	114.1 \pm 2.7*	172.1*
F100-3	225.5 \pm 4.0	340.6 \pm 6.0
F80	195.5 \pm 1.1	295.3 \pm 1.7

* Sharp et al. (1972).

Table 5

Episome	No. of supE ⁺ /donor	lip ⁺ /supE ⁺
F100-1	1.3×10^{-3}	8/85
F100-2 (F450)	5.8×10^{-4}	84/84
F152	1.2×10^{-3}	—
F152-1	4.9×10^{-5}	40/40
F153	1.0×10^{-3}	4/40
F8	8.6×10^{-7}	12/16*

* Four of the trp⁺ bacteria were found to be revertants of trp amber mutation. Therefore the actual ratio is 12/12

References

- Adler, J., and Kaiser, A. D. (1963). Virology, 19, 117-126.
- Adhya, S., Cleary, P., and Campbell, A. (1968). Proc. Nat. Acad. Sci., Wash. 61, 956-962.
- Bachmann, B. J. (1972). Bacteriol. Rev. 36, 525-557.
- Beckwith, J., R., Signer, E. R., and Epstein, W. (1966). Cold Spr. Harb. Symp. Quant. Biol. 31, 393-401.
- Cairns, J. (1963). Cold Spr. Harb. Symp. Quant. Biol. 28, 43-46.
- Cox, G. B., Gibson, F., Luke, R. K. J., Newton, N. A., O'Brien, I. B., and Rosenberg, H. (1970). J. Bacteriol. 104, 219-226.
- Davis, B. D., and Mingioli, E. S. (1950). J. Bacteriol. 60, 17-28.
- Epstein, W., and Davies, M. (1970). J. Bacteriol. 101, 836-843.
- Herbert, A. A., and Guest, J. R. (1968). J. Gen. Microbiol. 53, 363-381.
- Hirota, Y., and Sneath, P. H. A. (1961). Japan J. Genet. 36, 307-318.
- Hirsch, H. J., Saedler, H., and Starlinger, P. (1972). Mol. Gen. Genet. 115, 266-276.
- Hsu, M. T., and Davidson, N. (1972). Proc. Nat. Acad. Sci., Wash. 69, 2823-2827.
- Kim, J. S., and Davidson, N. (1973). Virology, in press.
- Kupor, S. R., and Frankel, D. G. (1969). J. Bacteriol. 100, 1296-1301.
- Lederberg, J. (1949). Proc. Nat. Acad. Sci., Wash. 35, 178-184.
- Low, K. B., Gates, F., Goldstein, T., and Soll, D. (1971). J. Bacteriol. 108, 742-750.

- Low, K. B. (1972). Bacteriol. Rev. 36, 587-607.
- Nomura, M., and Witten, C. (1967). J. Bacteriol. 94, 1093-1111.
- H., Shimada, K., and Tomizawa, J. (1968). Mol. Gen. Genet. 101, 227-244.
- Rolfe, B., and Einsenberg, M. A. (1968). J. Bacteriol. 96, 515-524.
- Rothman, J. L. (1965). J. Mol. Biol. 12, 892-912.
- Sharp, P. A., Hsu, M. T. Ohtsubo, E., and Davidson, N. (1972).
J. Mol. Biol. 71, 471-497.
- Signer, E. R., Beckwith, J. R., and Brenner, S. (1965).
J. Mol. Biol. 14, 153-166.
- Stouthamer, A. H., de Haan, P. G., and Nijkamp, H. J. J. (1965).
Genet. Res. 6, 442-453.
- Taylor, A. L., and Trotter, C. D. (1967). Bacteriol. Rev. 31,
332-353.
- Taylor, A. L., and Trotter, C. D. (1972). Bacteriol. Rev. 36,
504-524.
- Wallace, B. J., and Pittard, J. (1967). J. Bacteriol. 93, 237-244.
- Young, I. G., McCann, L. M., Stroobant, P., and Gibson, F. (1971).
J. Bacteriol. 105, 769-778.

Legends to Figures

Fig. 1. Heteroduplex of F8-33 with F. The two insertions, \underline{c} and $\zeta\epsilon$ of F8-33 are used as position markers for mapping F and bacterial sequences in other F-prime factors. The bacterial sequence is shown as ^a/sawtooth line.

Fig. 2. Summary of the genetic and physical structures of F-prime factors studied. (a) The bacterial DNA of episomes. The deletions are indicated by the dotted lines. A small I-D loop of F100-3 is not shown in this figure. The top line is the coordinate system of F100, F152, their derivatives, and F80. The bottom line is the coordinate system of F8 and its derivatives. (b) The F DNA of the episomes. The sequence deleted is indicated by the dotted line.

Fig. 3. Heteroduplexes involving F152. The sawtooth line is bacterial DNA. Numbers are distances in kilobases from the selected origins of F and bacterial DNA. The coordinates for bacterial and F sequence are labeled with a B and F respectively. (a) F152/F heteroduplex. The electronmicrograph is shown in Plate I. (b) F152/F8-33 heteroduplex. The two insertion loops of F8-33 are indicated as \underline{c} and $\zeta\epsilon$. Sequences connected by the dotted line sometimes form base pairs and is seen mostly in heteroduplexes between incomplete strands.

Fig. 4. Structure of F152/F152-1 heteroduplex. The corresponding electron micrograph is shown in Plate II.

Fig. 5. Heterduplex between F152-1 and F8-33. Structure (b) is formed from (a) by the base pairing between sequences $\underline{f'g'}$ and \underline{fg} as indicated by the dotted line. It is seen in heteroduplexes between two incomplete strands. The micrograph corresponding to structure (a) is shown in Plate III.

Fig. 6. Heteroduplex of F100-1 with F152-1. An electron micrograph is shown in Plate IV.

Fig. 7. Heteroduplex of F100-1 with F152. Note that the sequence around 0 F/0 B is homologous. In the F100-1/F152-2 heteroduplex, a small deletion at 151.7B is observed (see Sec. D and Plate VIII).

Fig. 8. Structure of the F100-1/F100-2 heteroduplex. Plate V is the corresponding micrograph.

Fig. 9. Structure of the F100-2/F152-1 heteroduplex. Note that the deletion end points of F100-2 is in the single-strand loop leaving the duplex at point a. See also Plate VI.

Fig. 10(a). Heteroduplex between F100-1 and F8-33. The sequences connected by the dotted line sometimes form base pairs. (b) F100-1/F8 heteroduplex.

Fig. 11. The structure of F8-3/F8-33 heteroduplex.

Fig. 12. Reconstruction of F152 or F152-2 episome from F152-1. In the case of F152-2 reconstruction, the bacterial chromosome carries a deletion in the kdp sequences.

Fig. 13. Reconstruction of F100 and analysis of supE marker.

Fig. 14. The structure of the F100-3/F152-2 heteroduplex. There is a small I-D loop of 0.7 kb at point c.

Fig. 15. Proposed mechanism for the formation of F80 from F8.

Fig. 16. Heteroduplexes involving F80. (a) The F80/F heteroduplex. (b) The F80/F8-33 heteroduplex. (c) F80/F100-1 heteroduplex. Electron micrographs of (b) and (c) are shown in Plates X and XI respectively.

Fig. 17. Probable processes for the formation of F100 and F152. Note that the excision point on F is the same for both episomes.

Legend to the Plates

Plate I. An F152/F heteroduplex. The schematic representation is shown in Fig. 3a. Letters a and b points to the junctions of duplex and single-stranded DNA at (0 F/0 B) and (2.8F/218.4B) respectively. Duplex F DNA, single-stranded F DNA and single-stranded bacterial DNA are identified as fd, fs, and bs respectively. Several ϕ X single-strand DNA are also marked.

Plate II. Electron micrograph of an F152/F152-1 heteroduplex. The arrow marks the point where the single-stranded loop, bs, comes out from the duplex region, bfd.

Plate III. Structure of an F152-1/F8-33 heteroduplex. The structures is illustrated in Fig. 5a. The junctions of the substitution loop are marked by the letters a and b. The letter e points to the place where the I-D loop, bfs, comes out of the duplex region. The two insertion loops of F8-33 are labeled by c and $\zeta\epsilon$. bfs indicates that the single-strand DNA is composed of both bacterial and F sequences.

Plate IV. The F100-1/F152-1 heteroduplex showing a substitution loop (ab) and an I-D loop (c). The schematic representation of this molecule is shown in Fig. 6.

Plate V. An electron micrograph of the F100-1/F100-2 heteroduplex. The substitution loop is identified by the two labels a and b.

Plate VI. Structure of the heteroduplex formed F100-2 and F152-1. The two single-stranded loops are identified by a and b, respectively.

Plate VII. The F100-1/F8 heteroduplex. The structure is illustrated in Fig. 10. The junctions between double-stranded DNA and single strand DNA are marked by a and b in the substitution loop and c in the deletion loop.

Plate VIII. Heteroduplex between F100-1 and F152-2 showing three I-D loops. I-D loop a is the sequence containing att^λ, bio, and uvrB. I-D loop b is the bacterial sequence deleted in F100-1 and I-D loop c the kdp deletion of F152-2.

Plate IX. An electron micrograph of the F100-2/F152-2 heteroduplex. Three I-D loops at a, b, and c can be seen. The I-D loop c is very small and appears to be a small stem in this magnification.

Plate X. The F80/F8-33 heteroduplex. The structure is illustrated in Fig. 17b.

Plate XI. Structure of the F80/F100-1 heteroduplex showing two I-D loops. The larger one is the bacterial sequence deleted in F100-1.

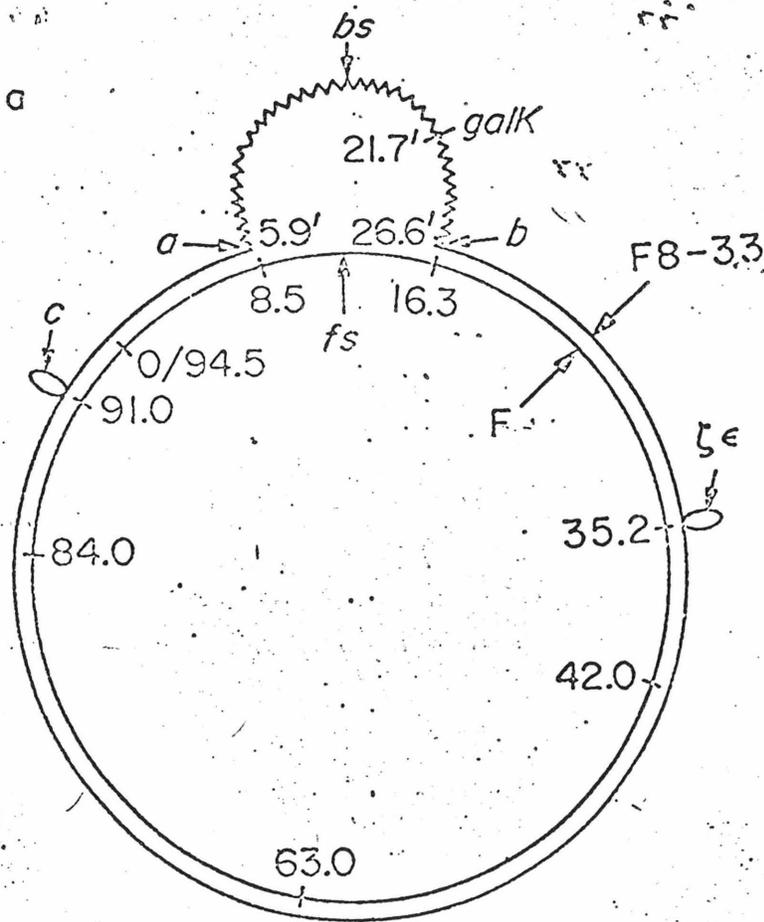
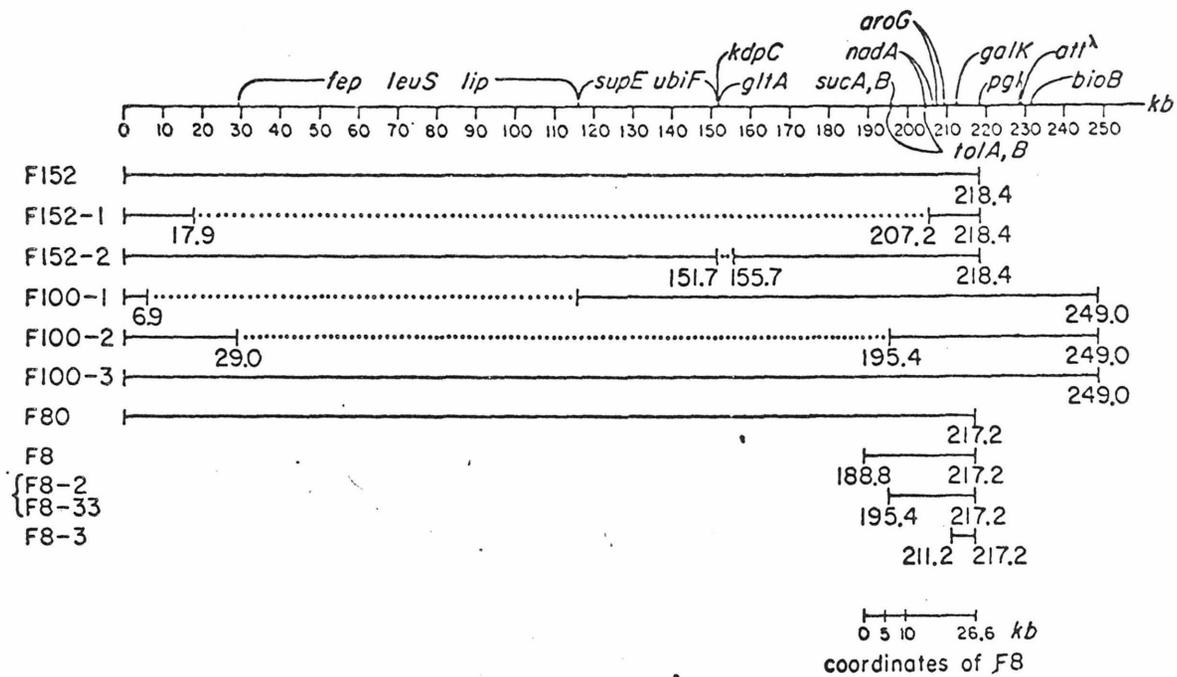
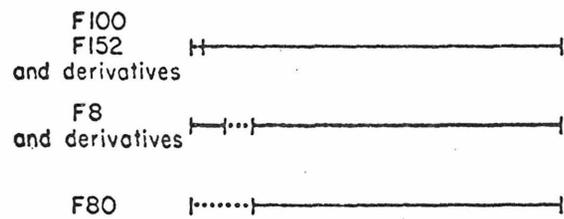


Fig. 1



(a)



(b)

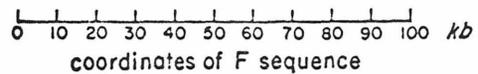


Fig. 2

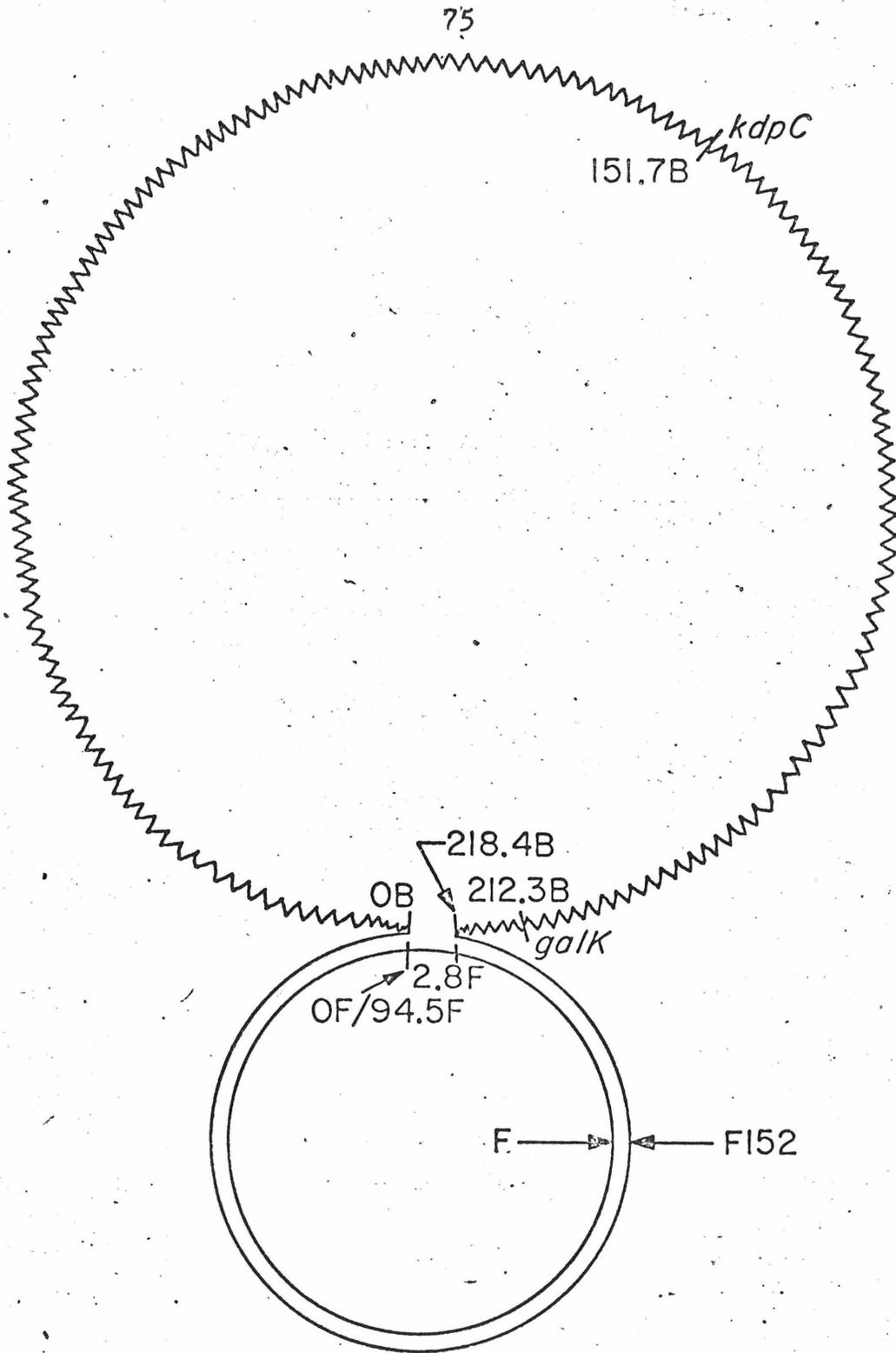


Fig. 3a

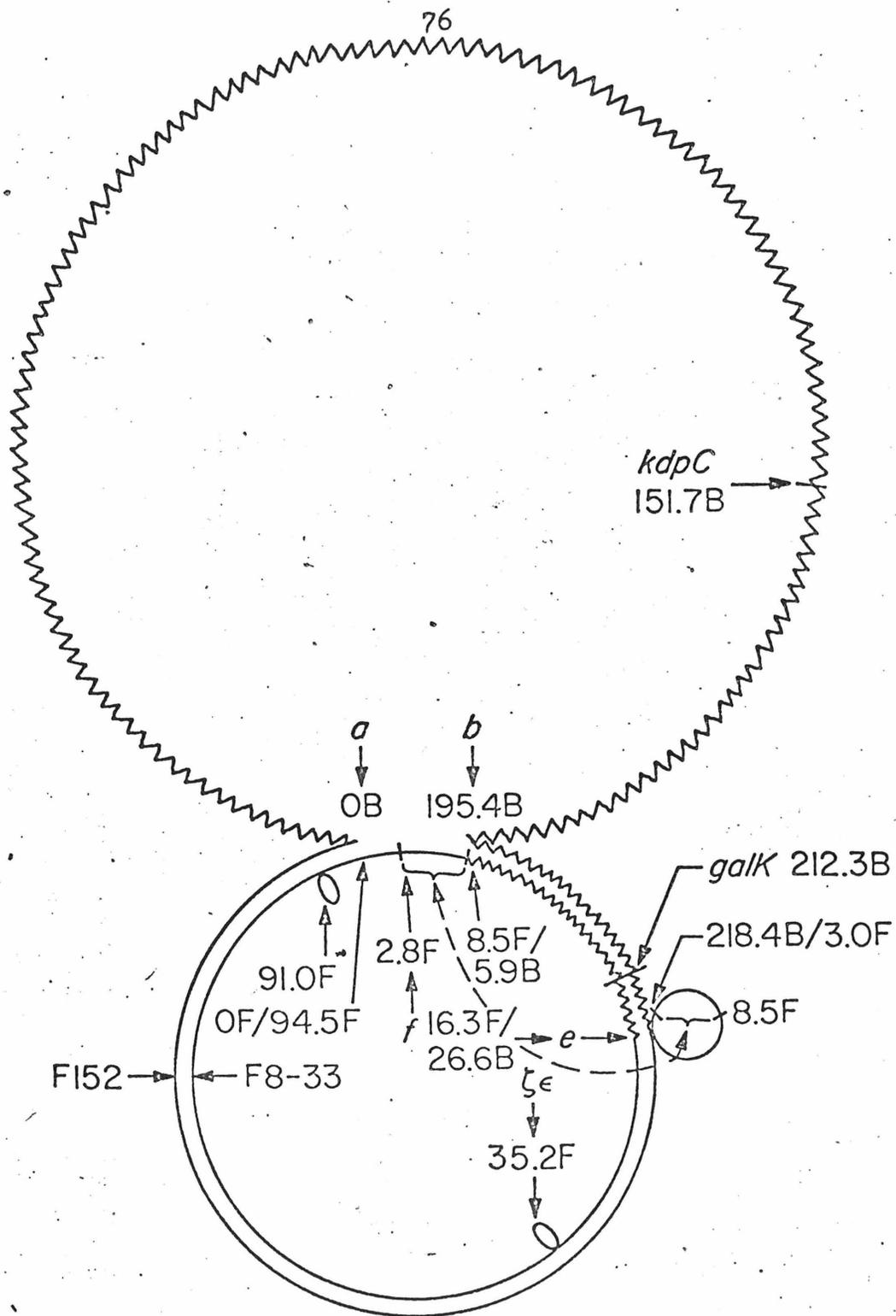


Fig. 3b

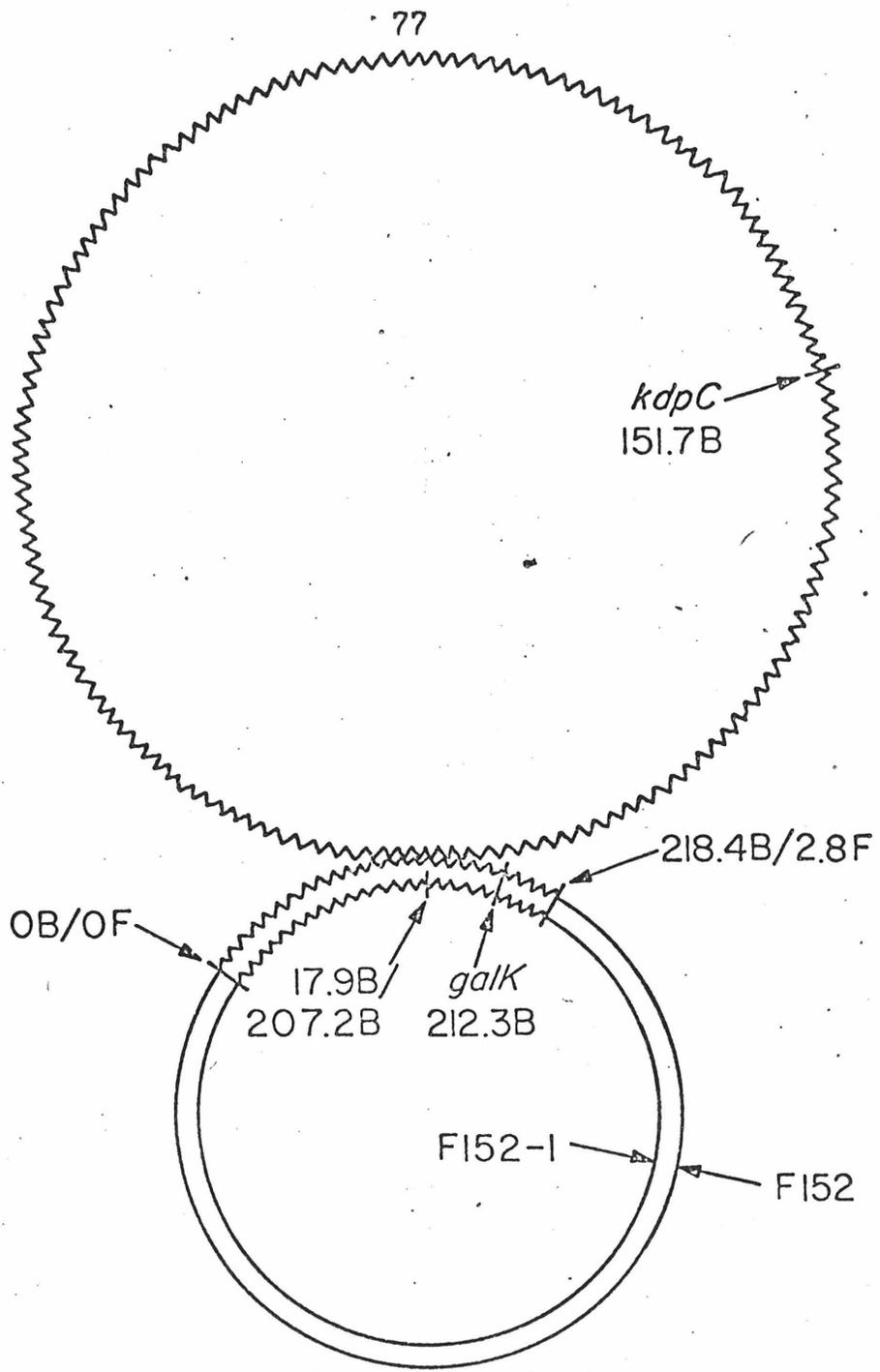


Fig. 4

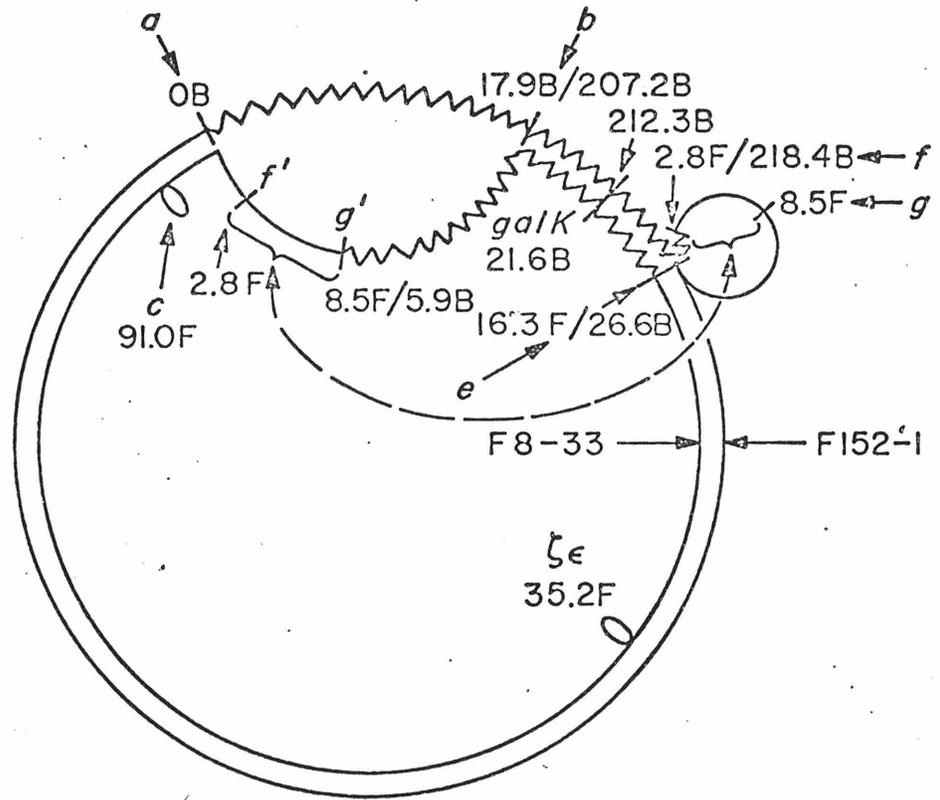


Fig. 5a

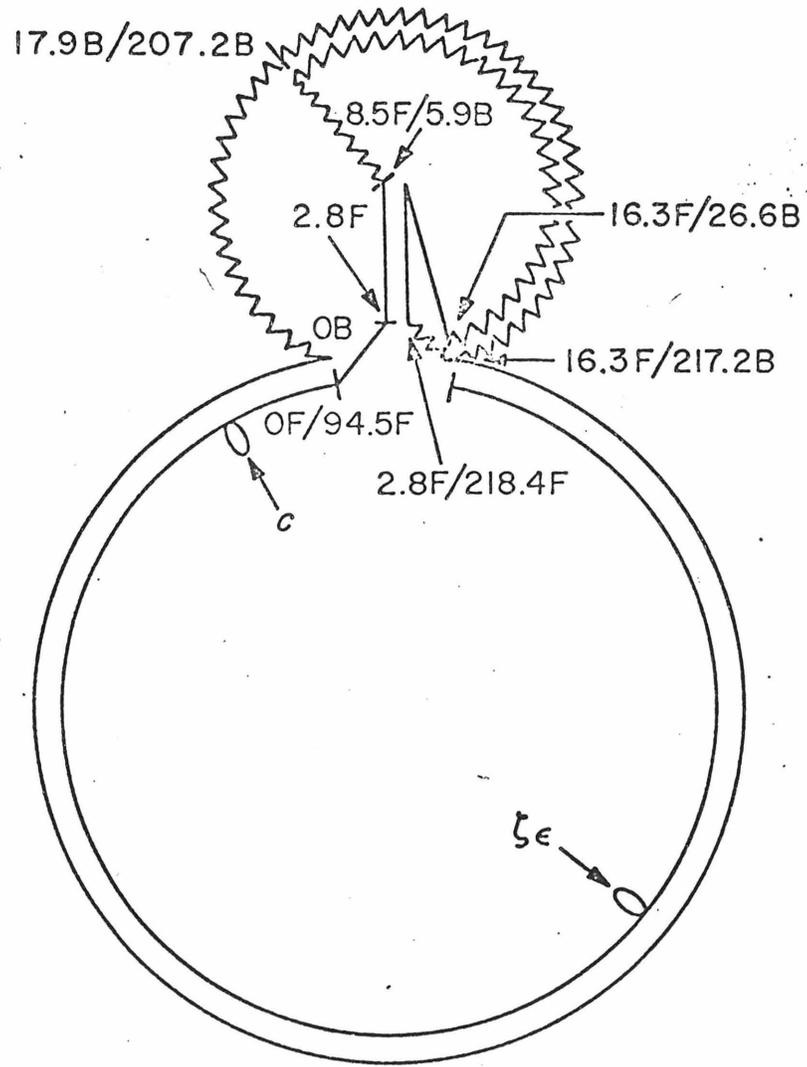


Fig. 5b

56

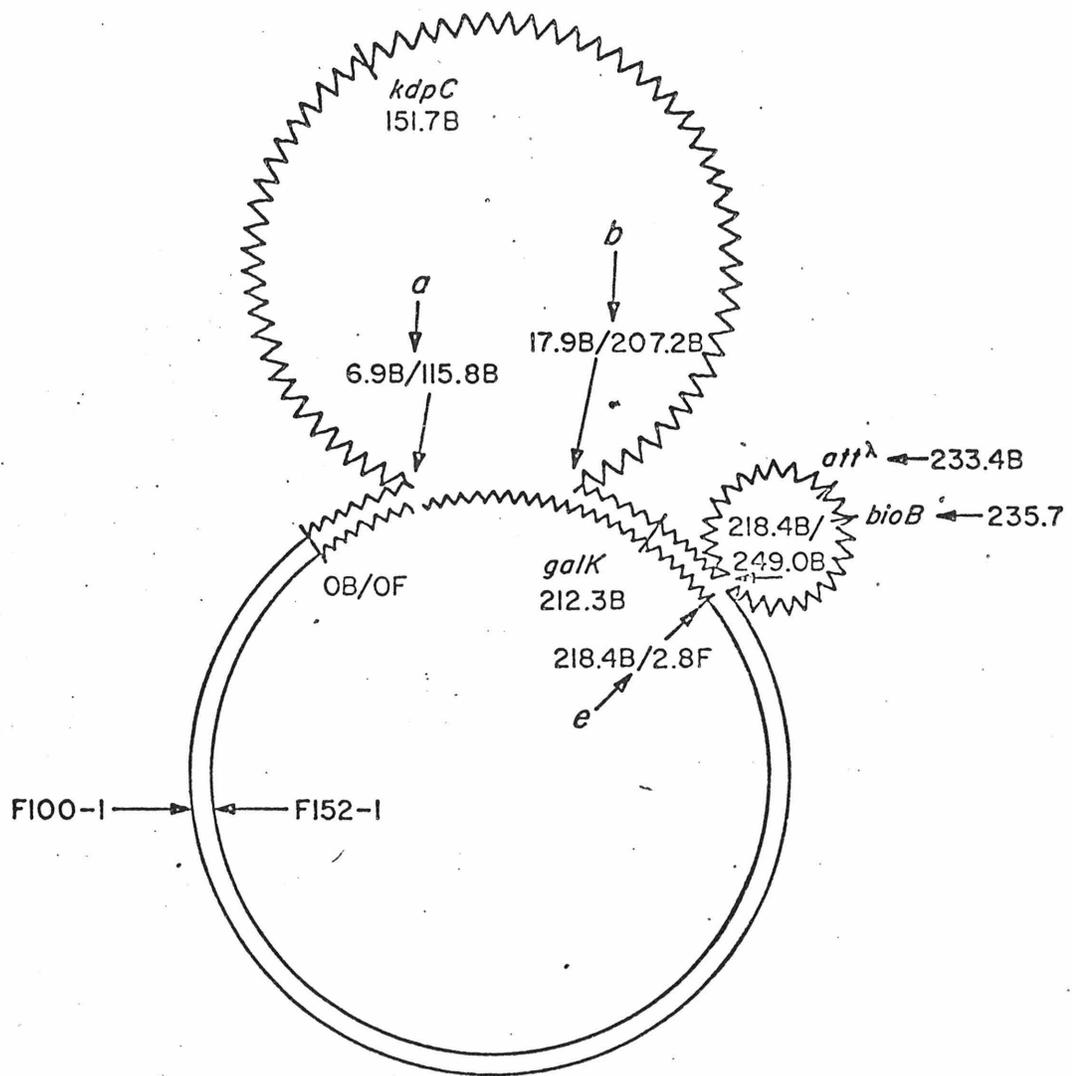


Fig. 6

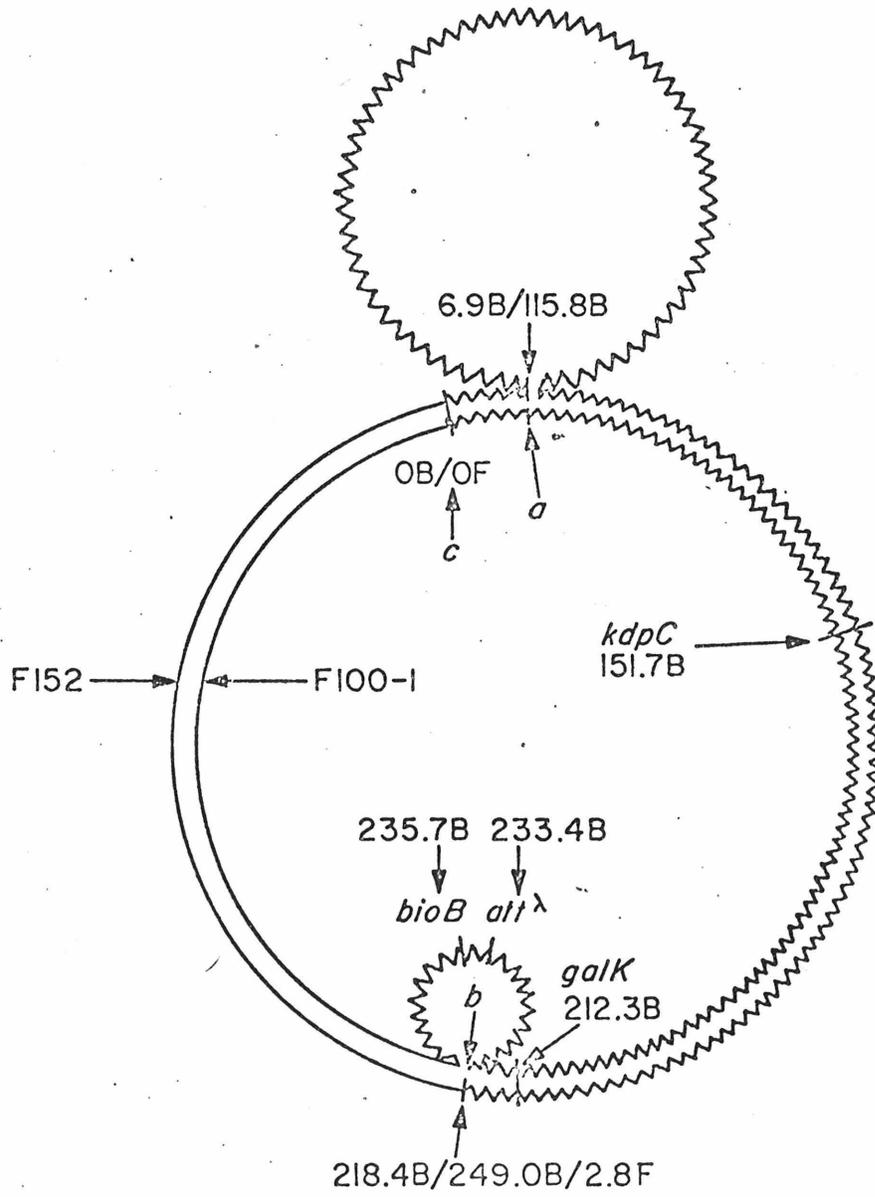


Fig. 7

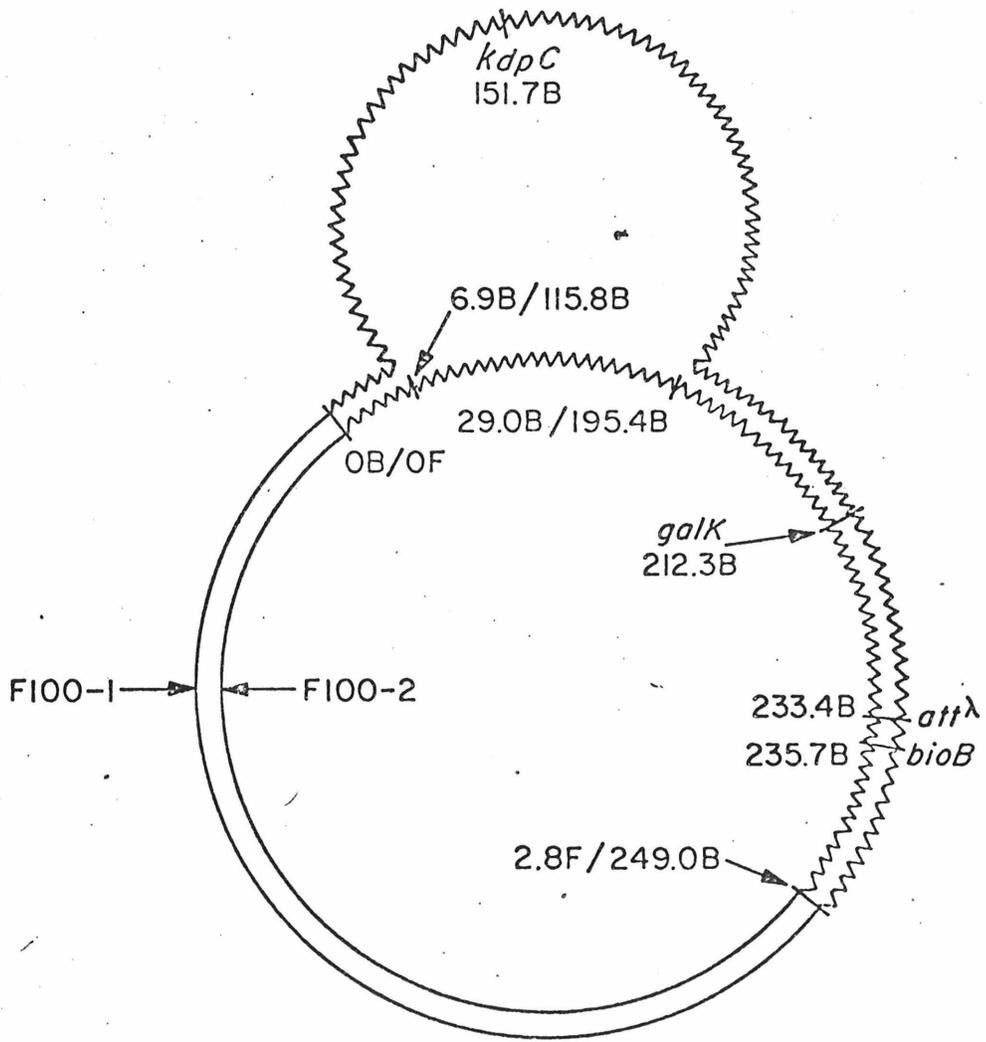


Fig. 8

8

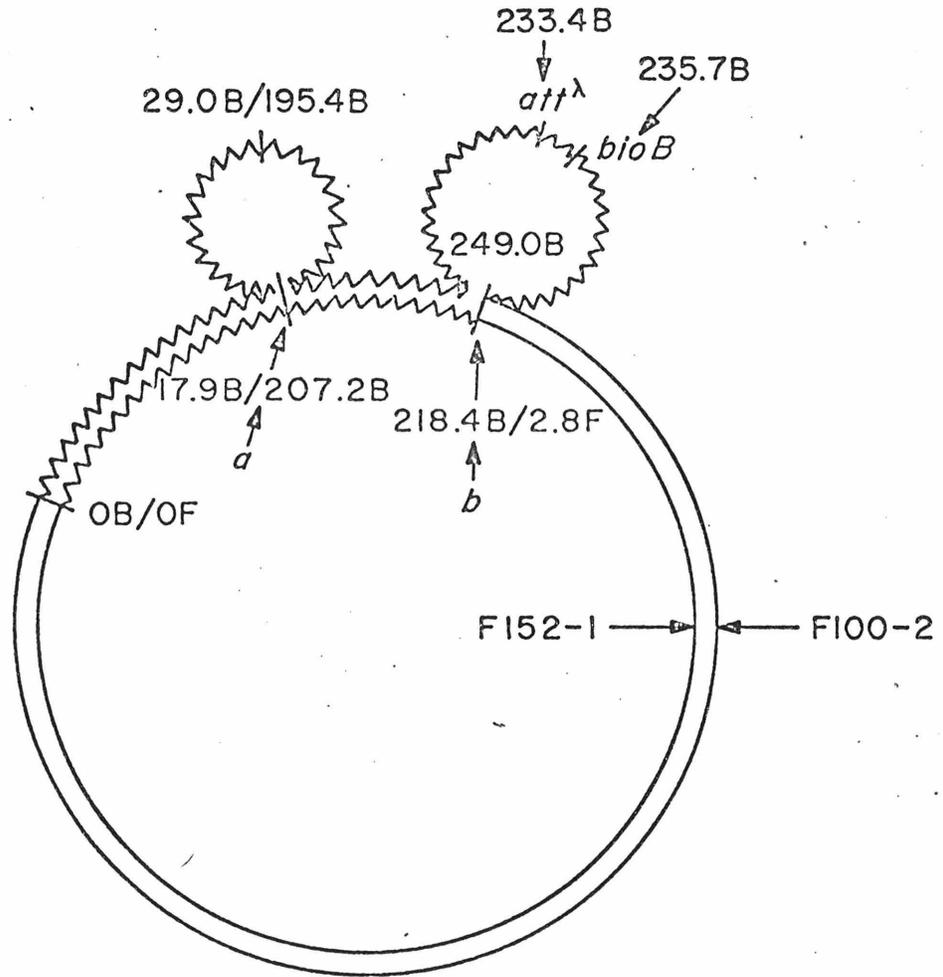


Fig. 9

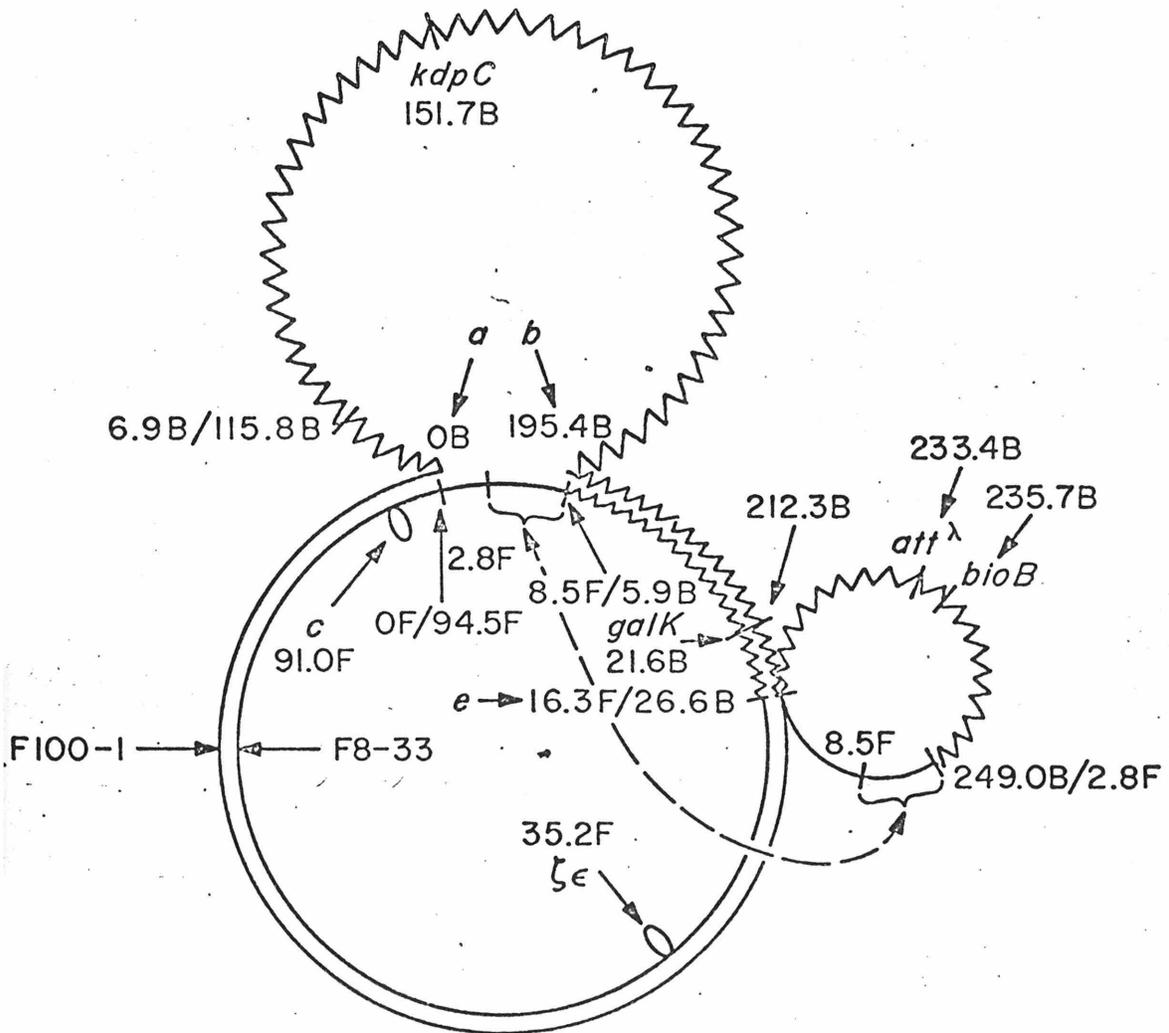


Fig. 10a

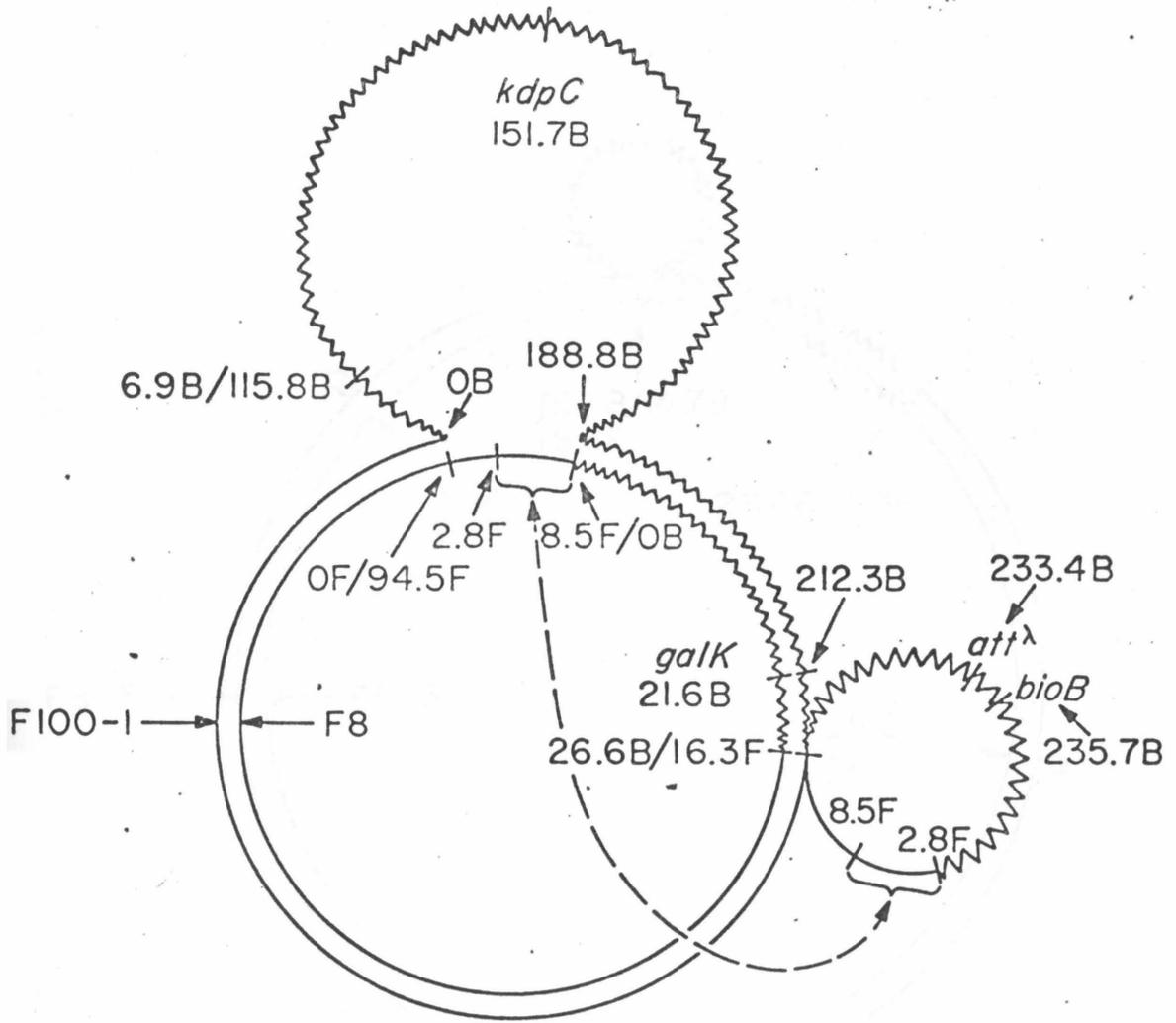


Fig. 10b

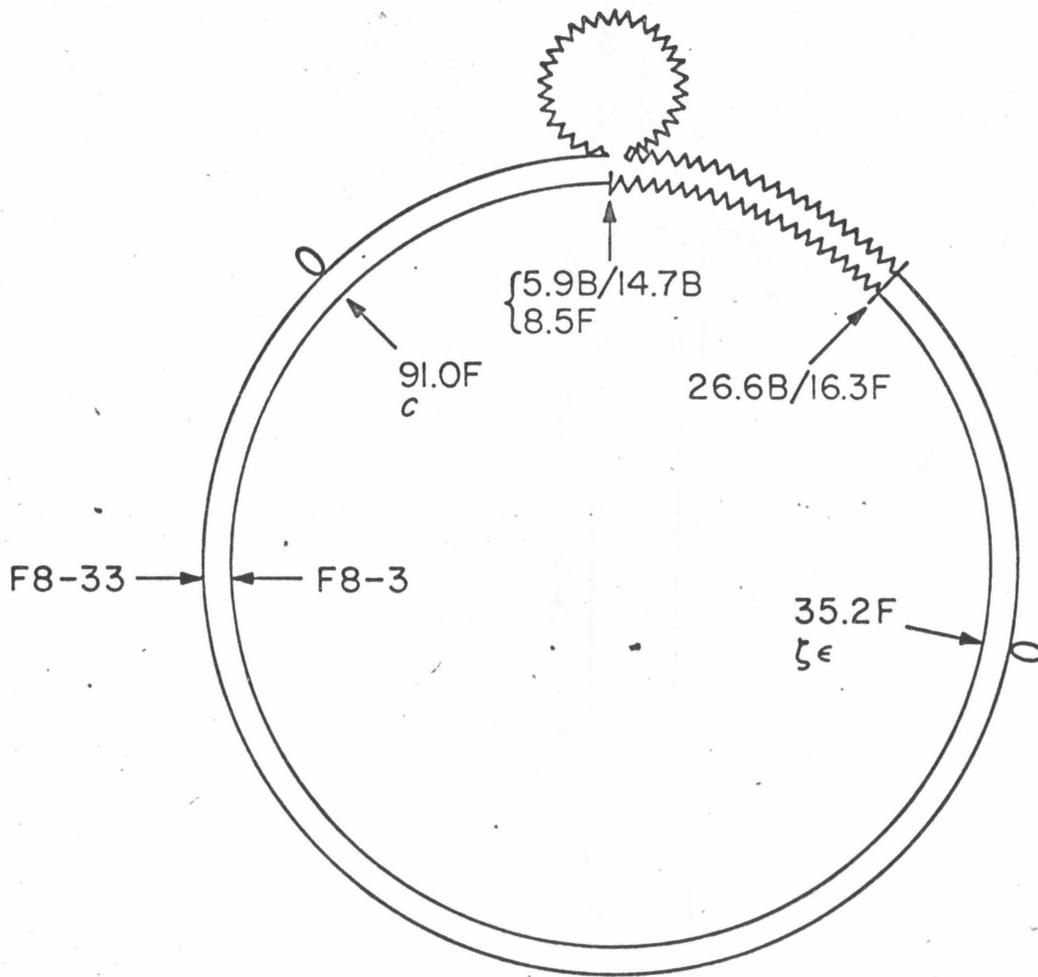


Fig. 11

21

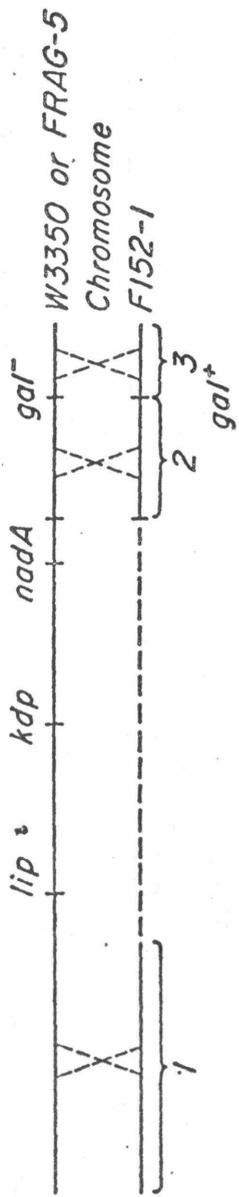


Fig. 12

13

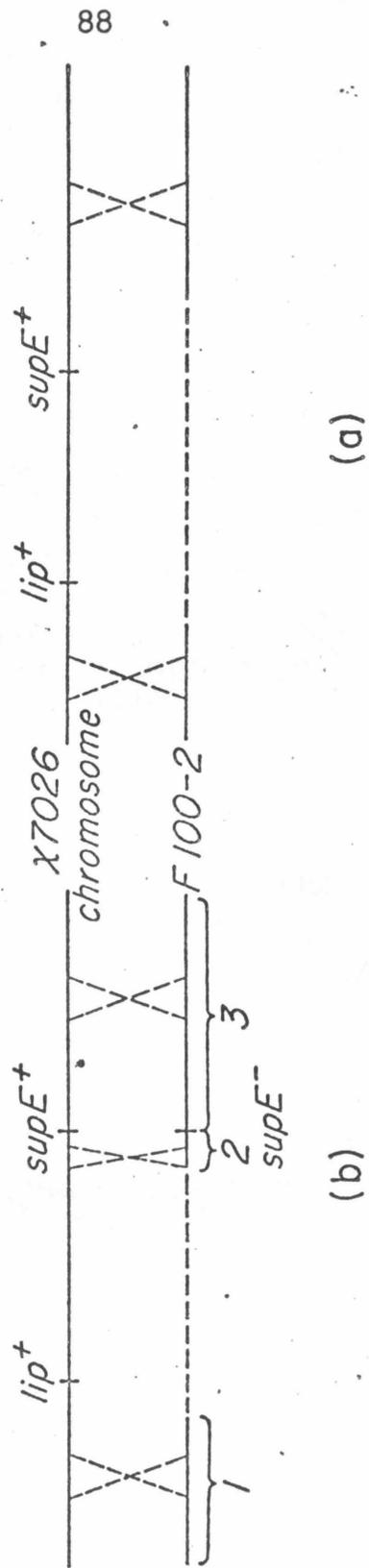


Fig. 13

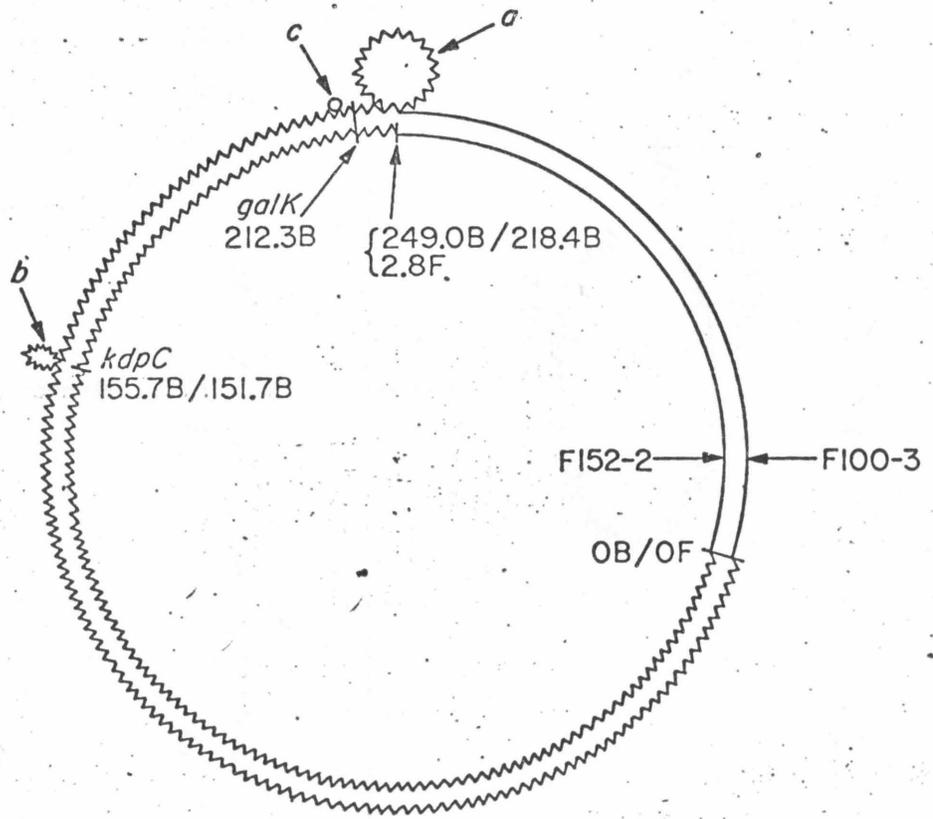


Fig. 14

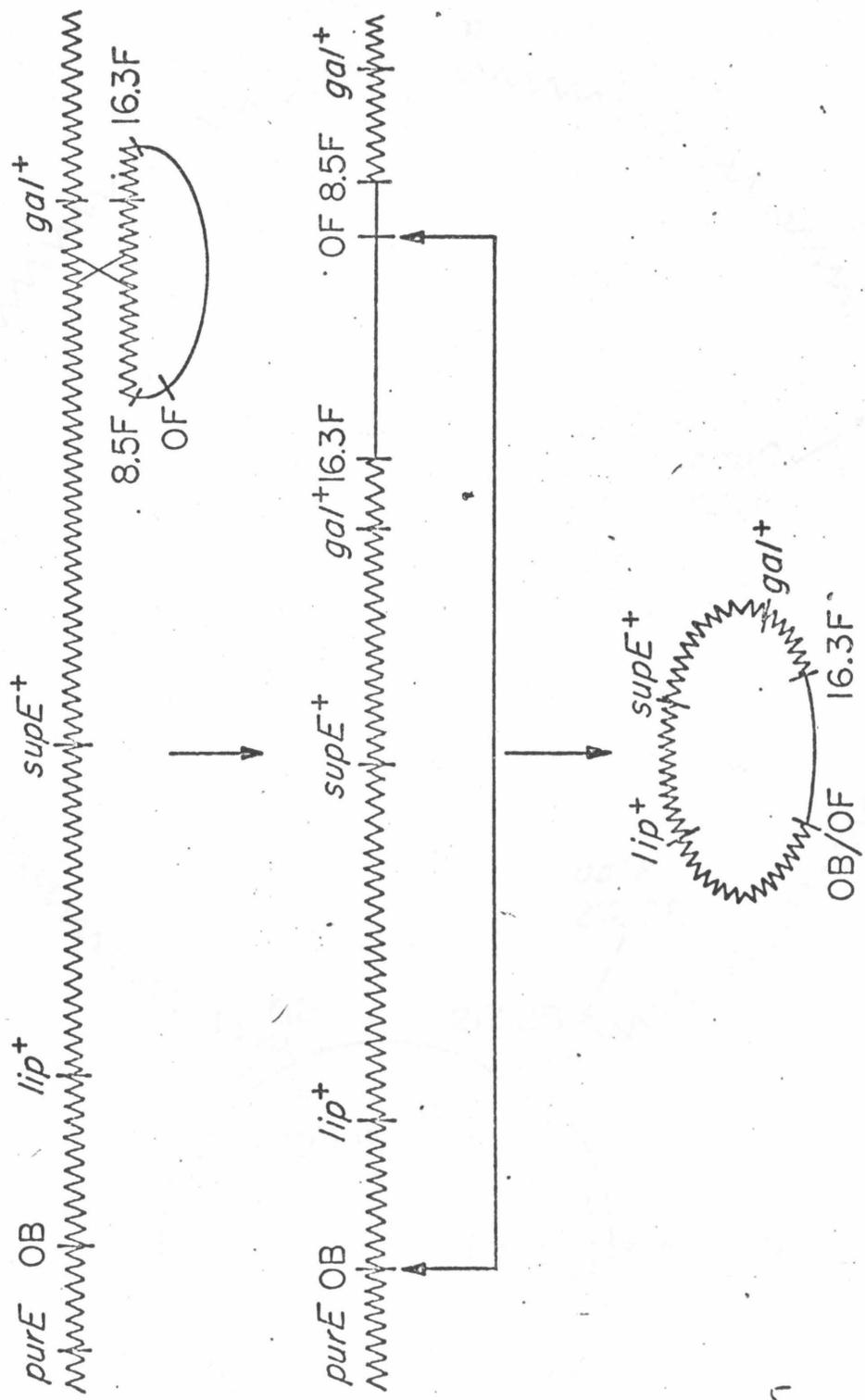


Fig. 15

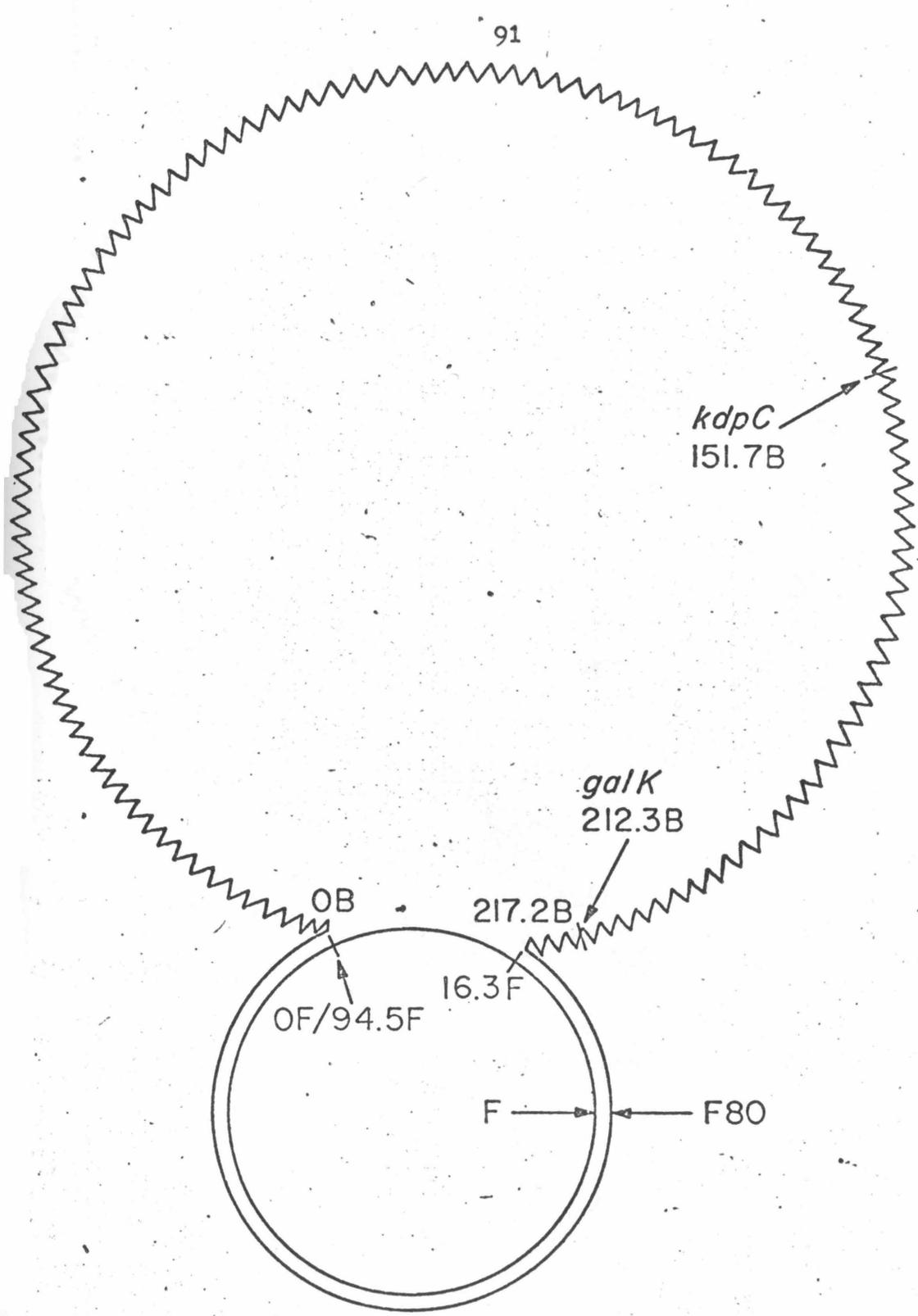


Fig. 16a

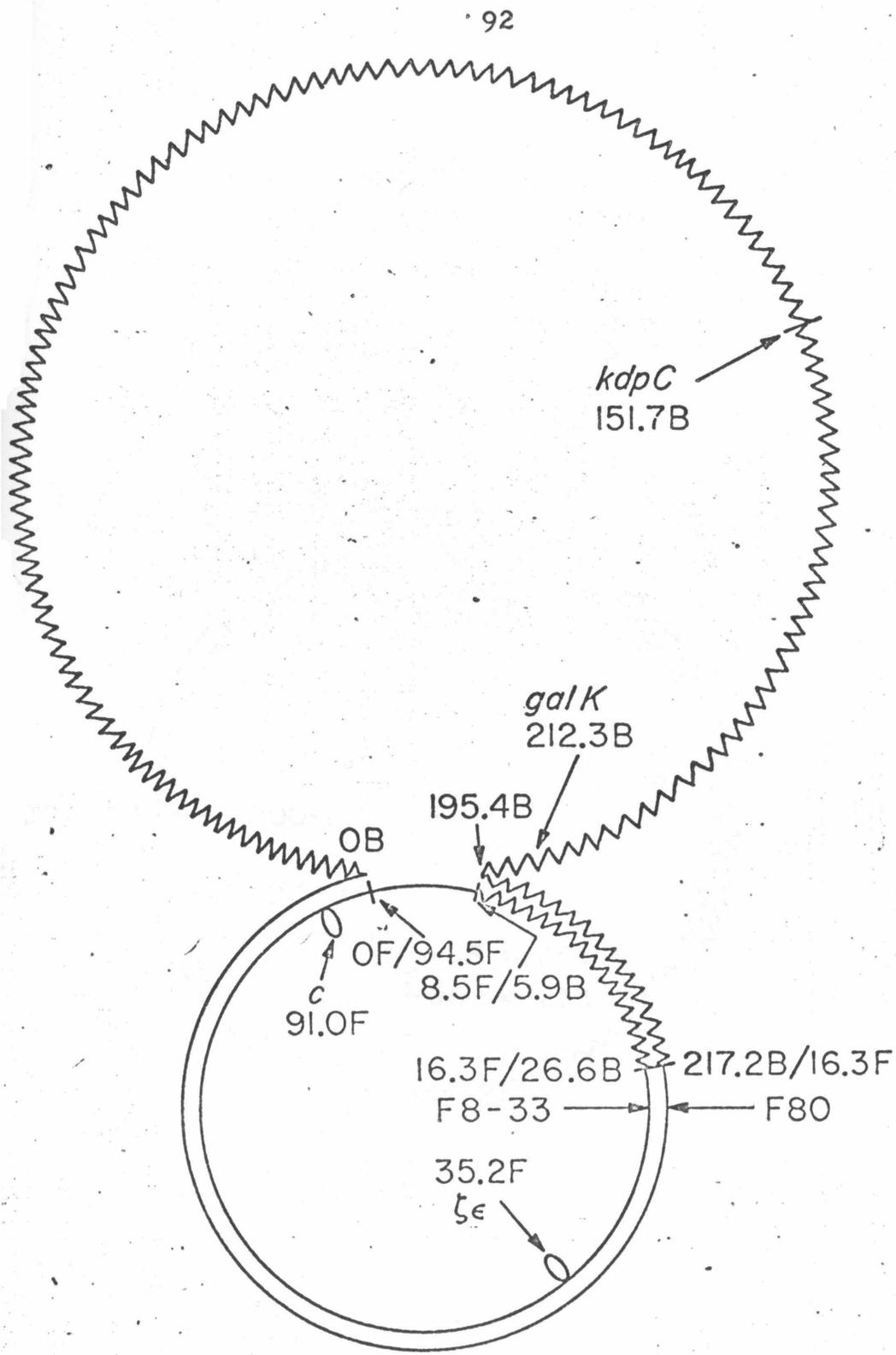


Fig. 16b

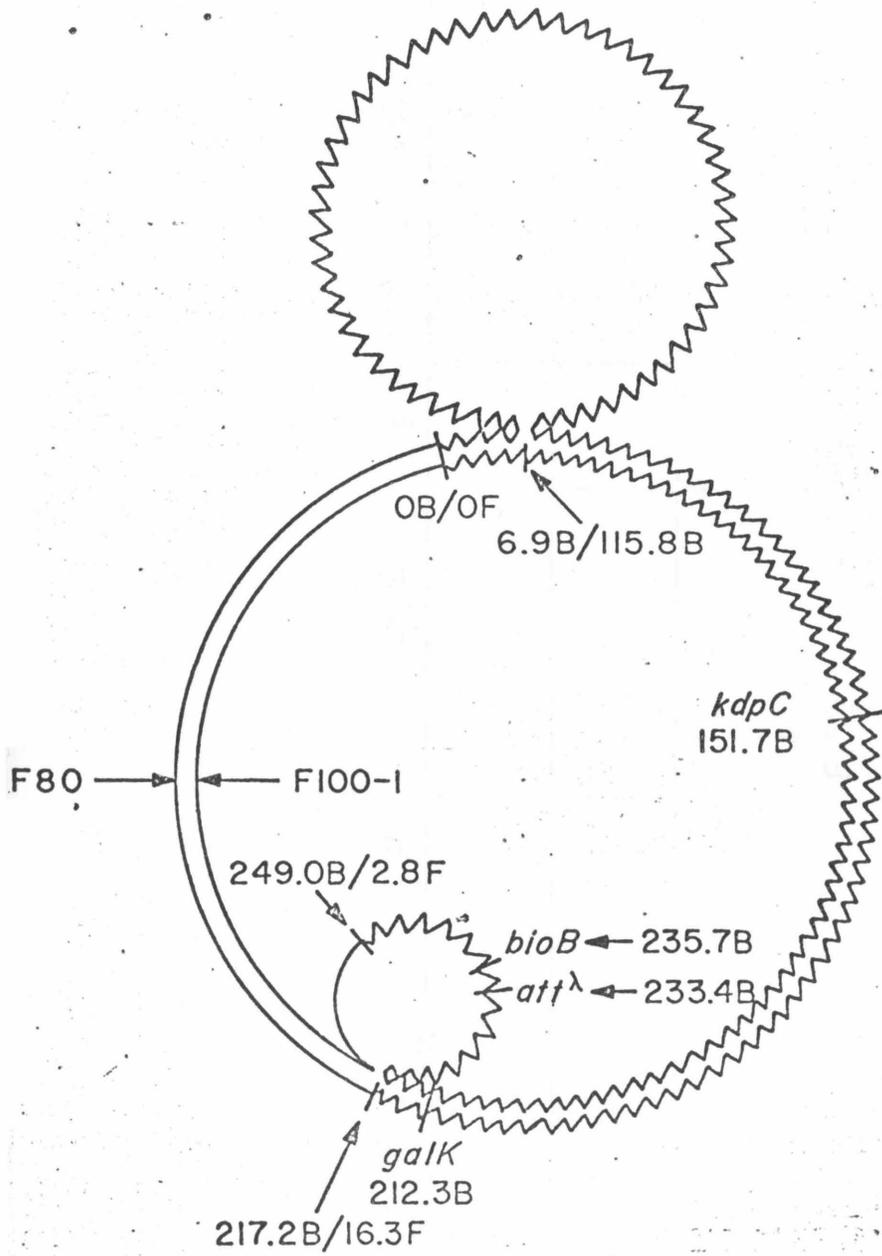


Fig. 16c

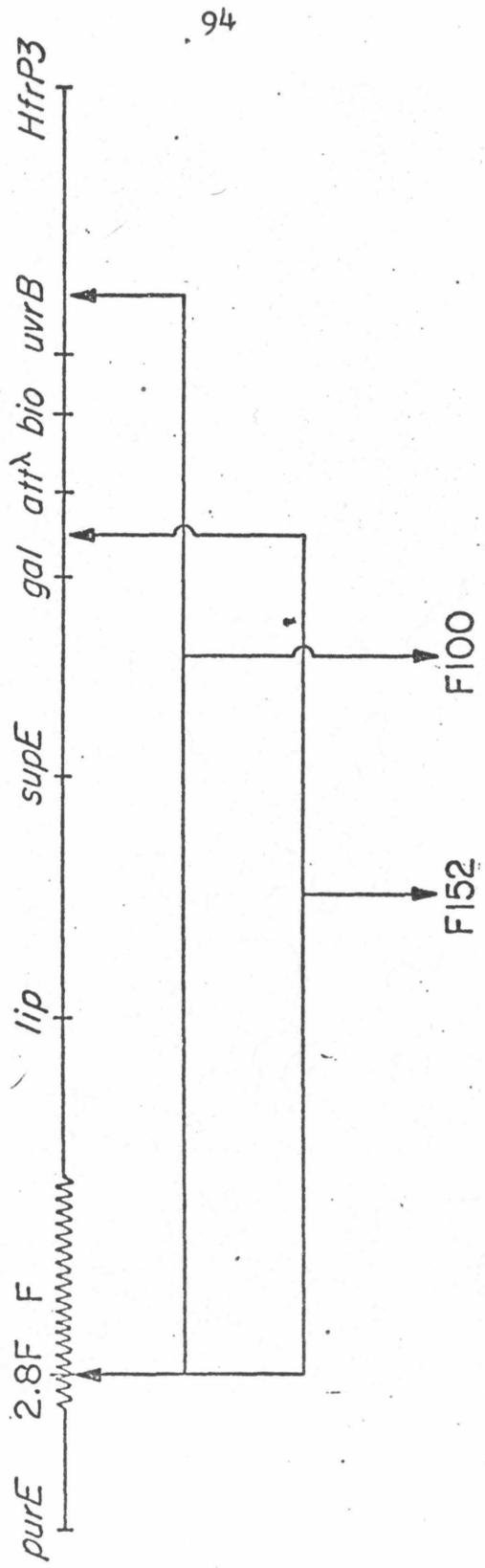


Fig. 17

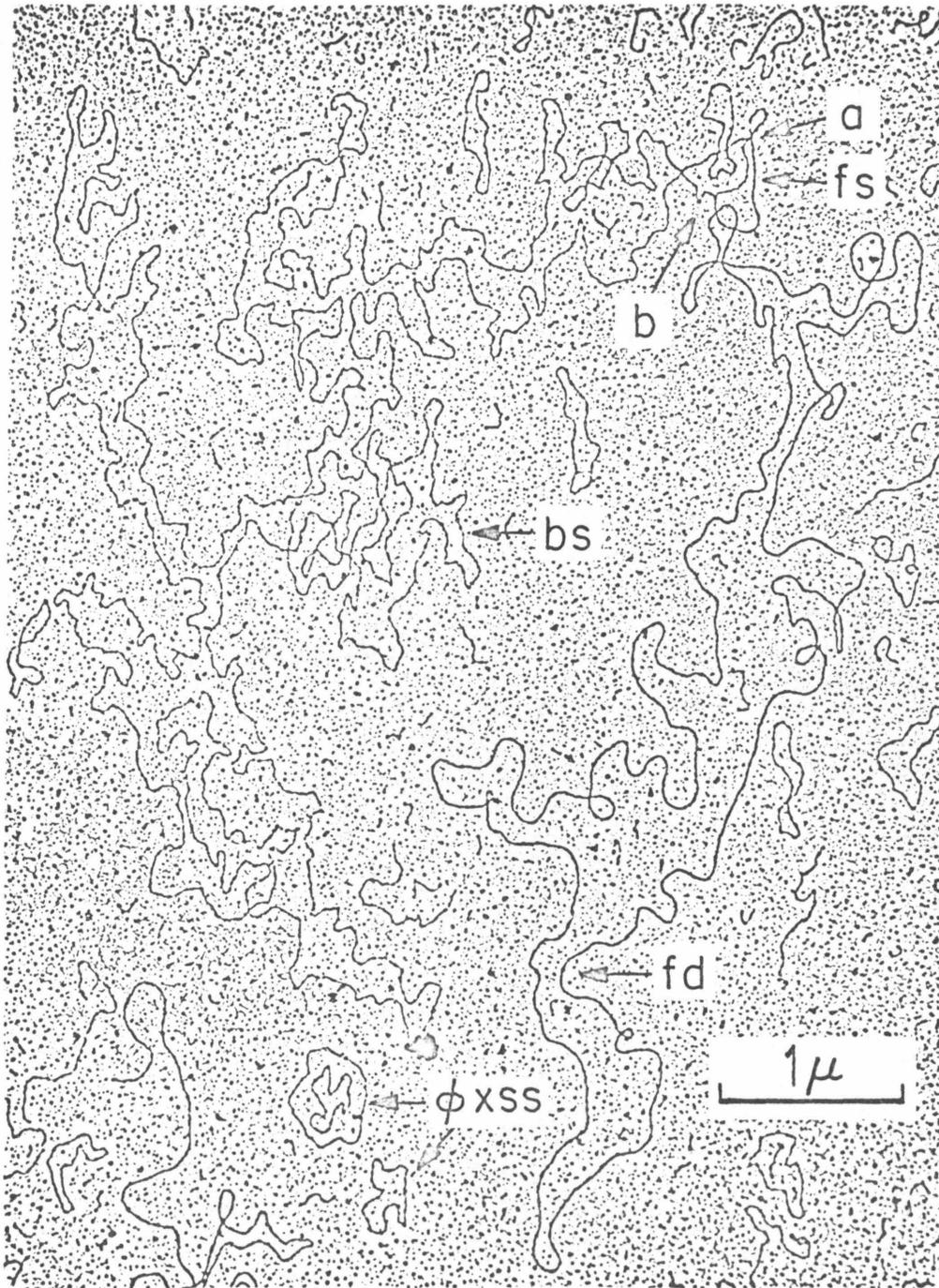


Plate I

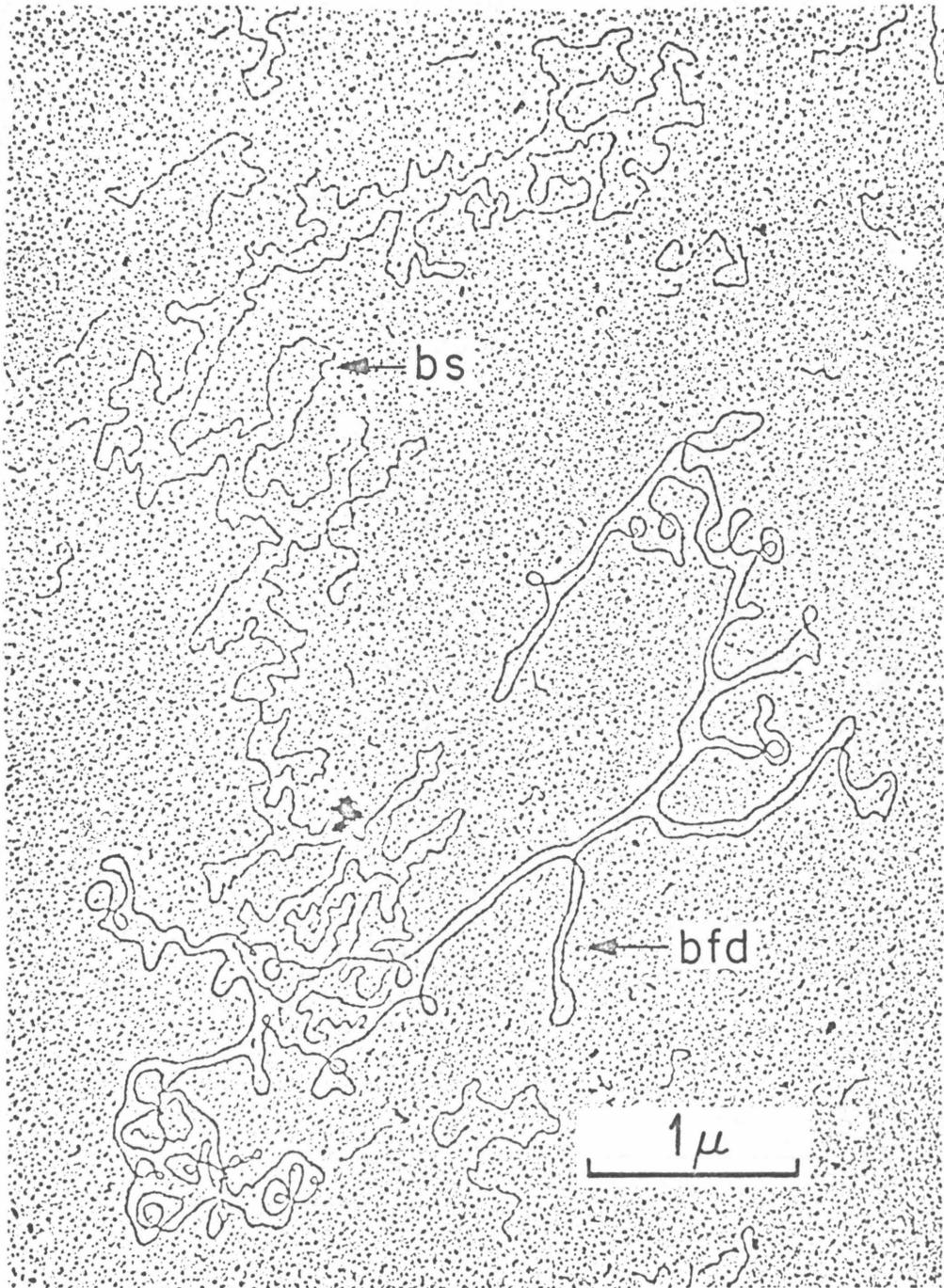


Plate II

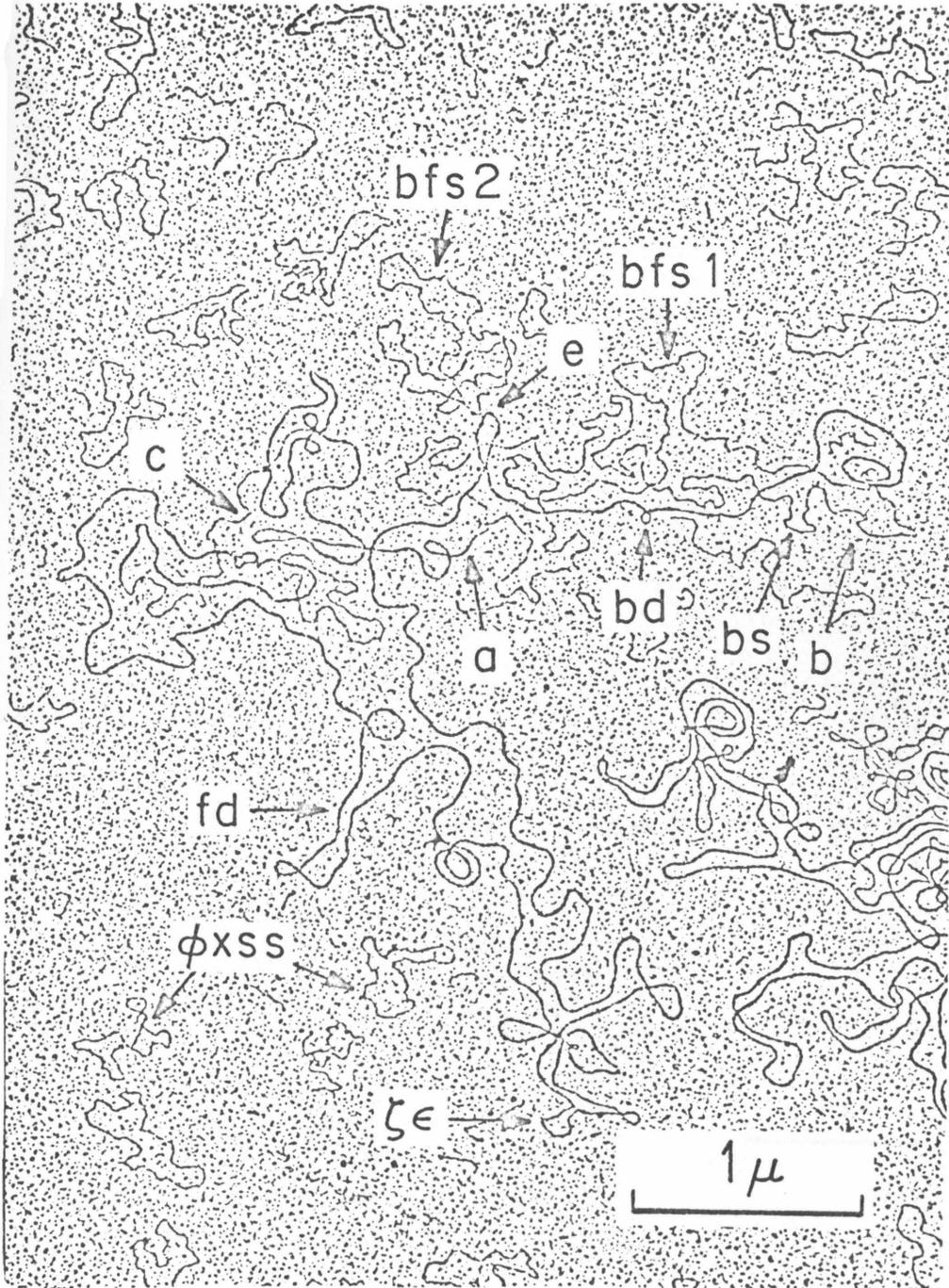


Plate III

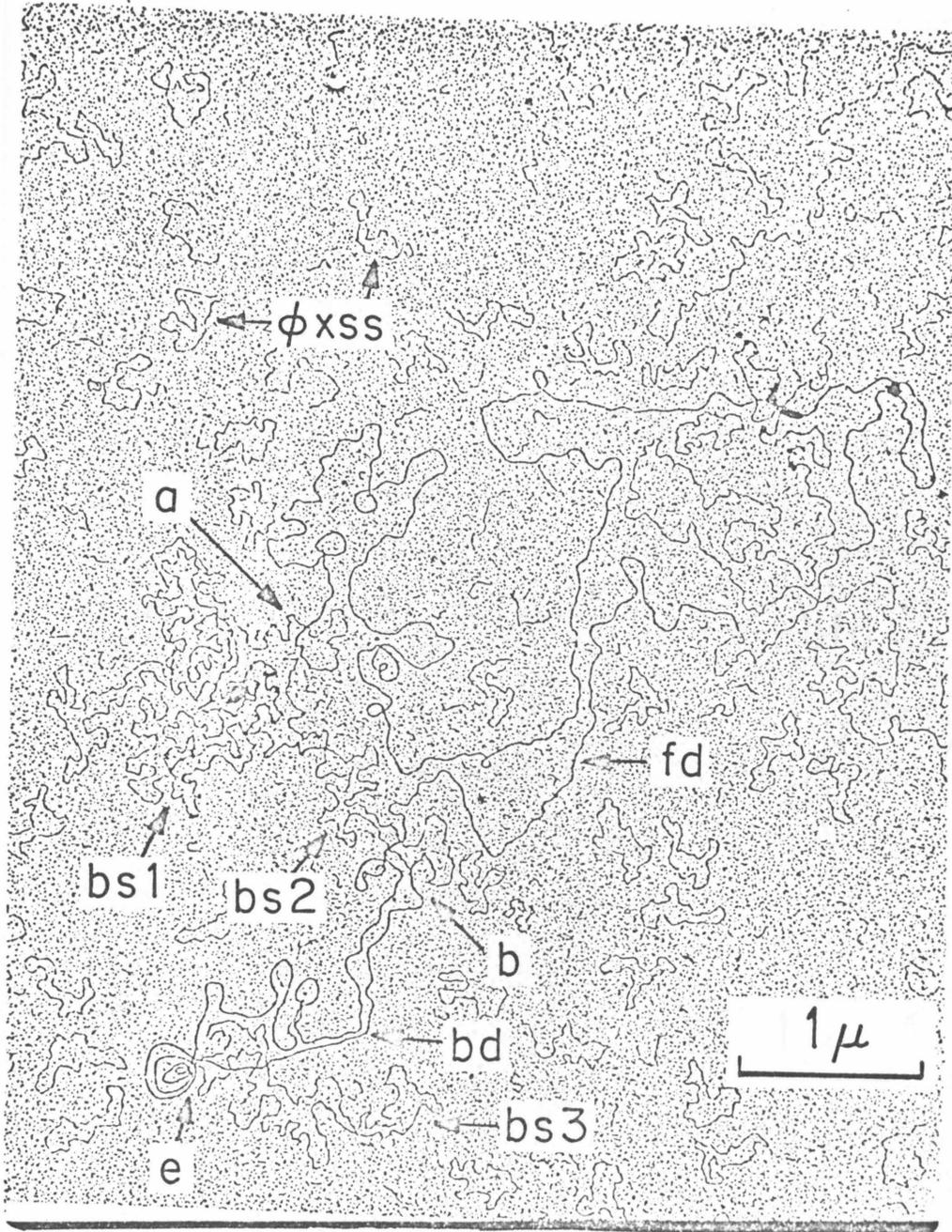


Plate IV

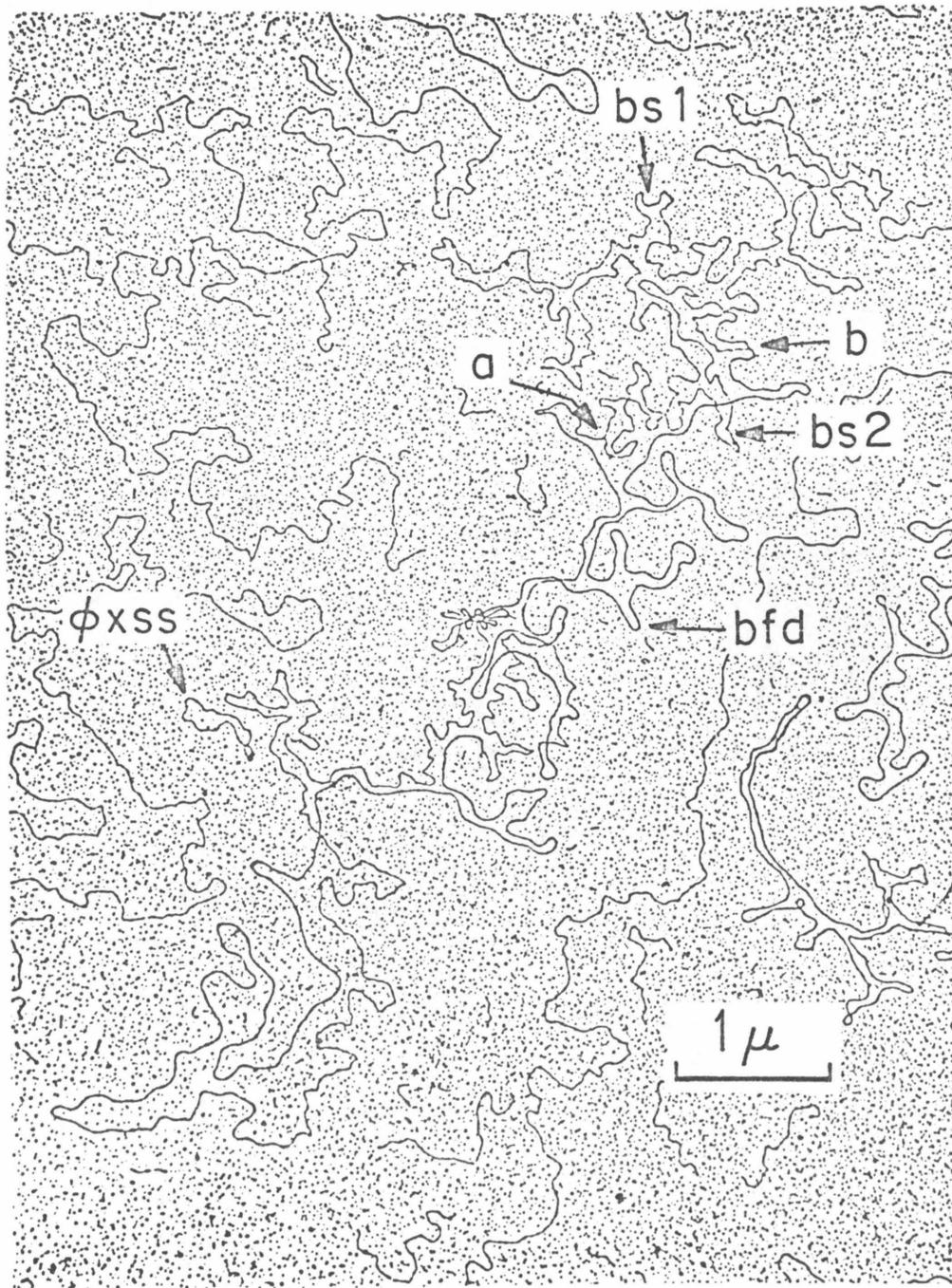


Plate V

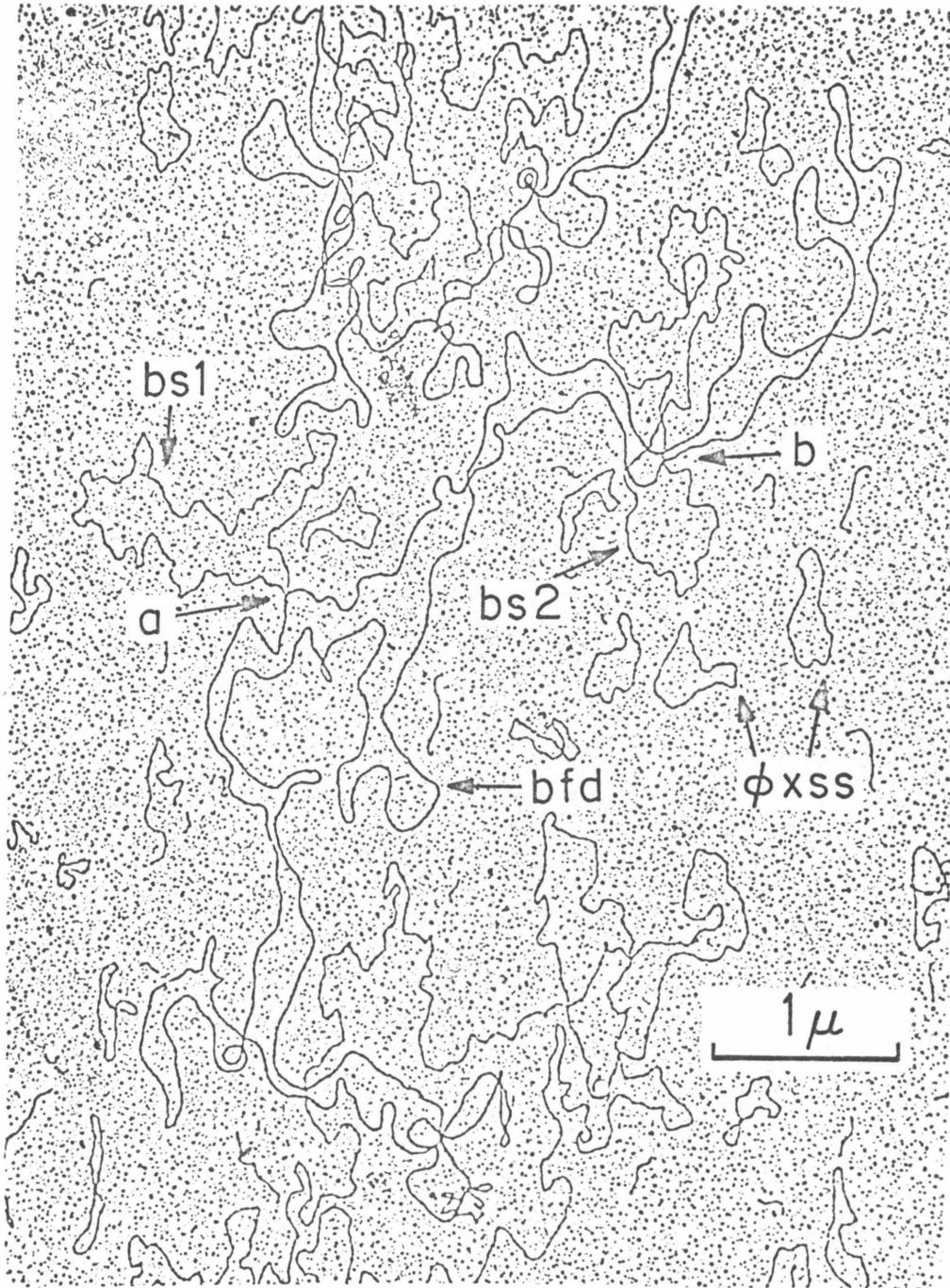


Plate VI

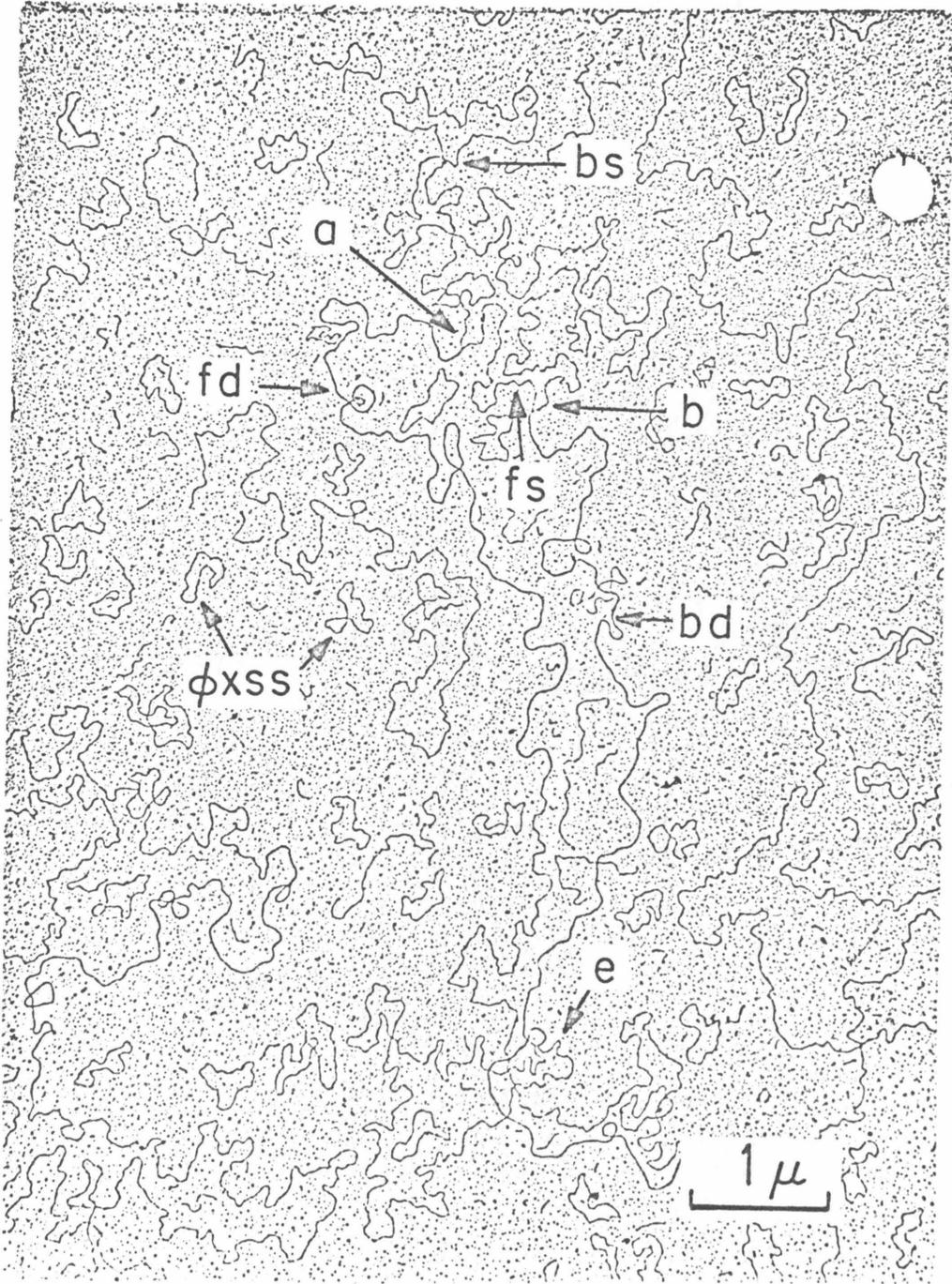


Plate VII

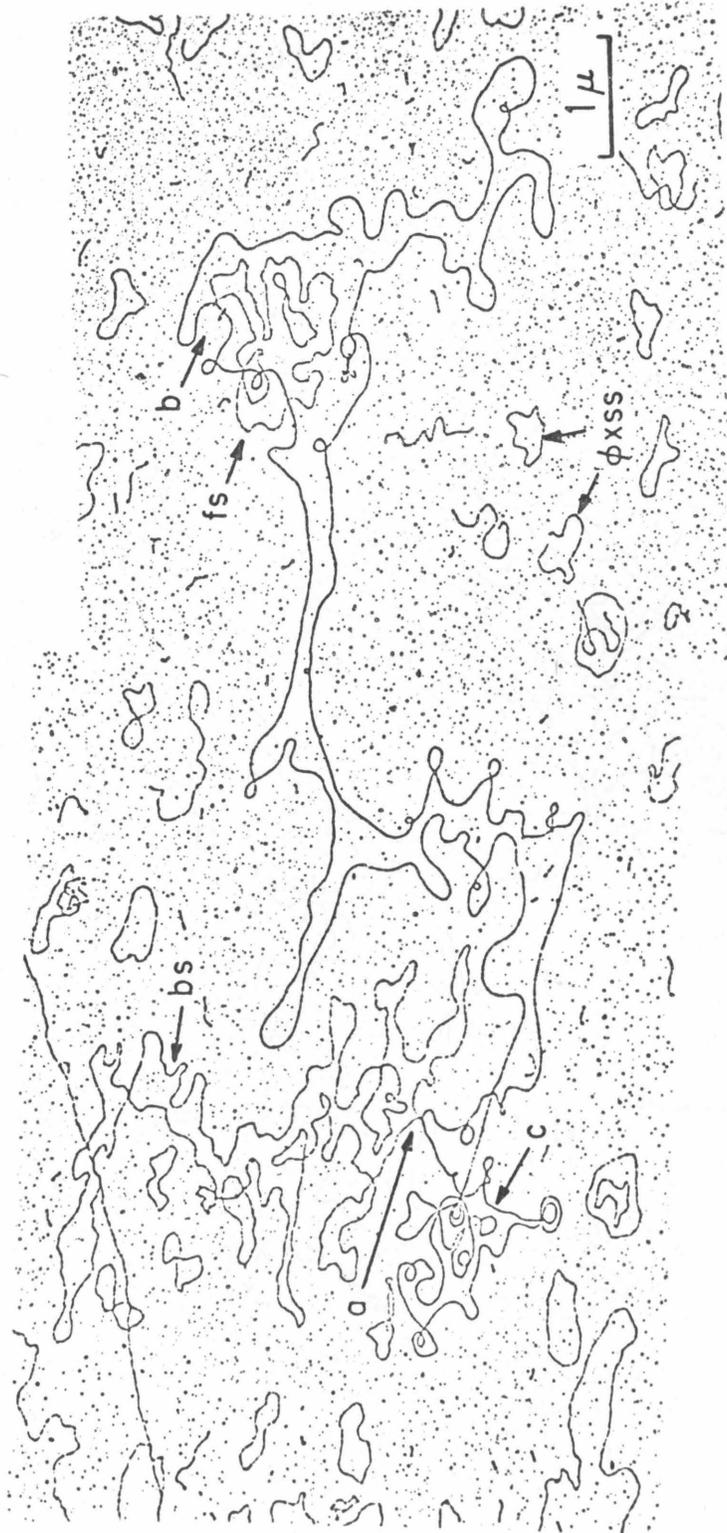


Plate VIII

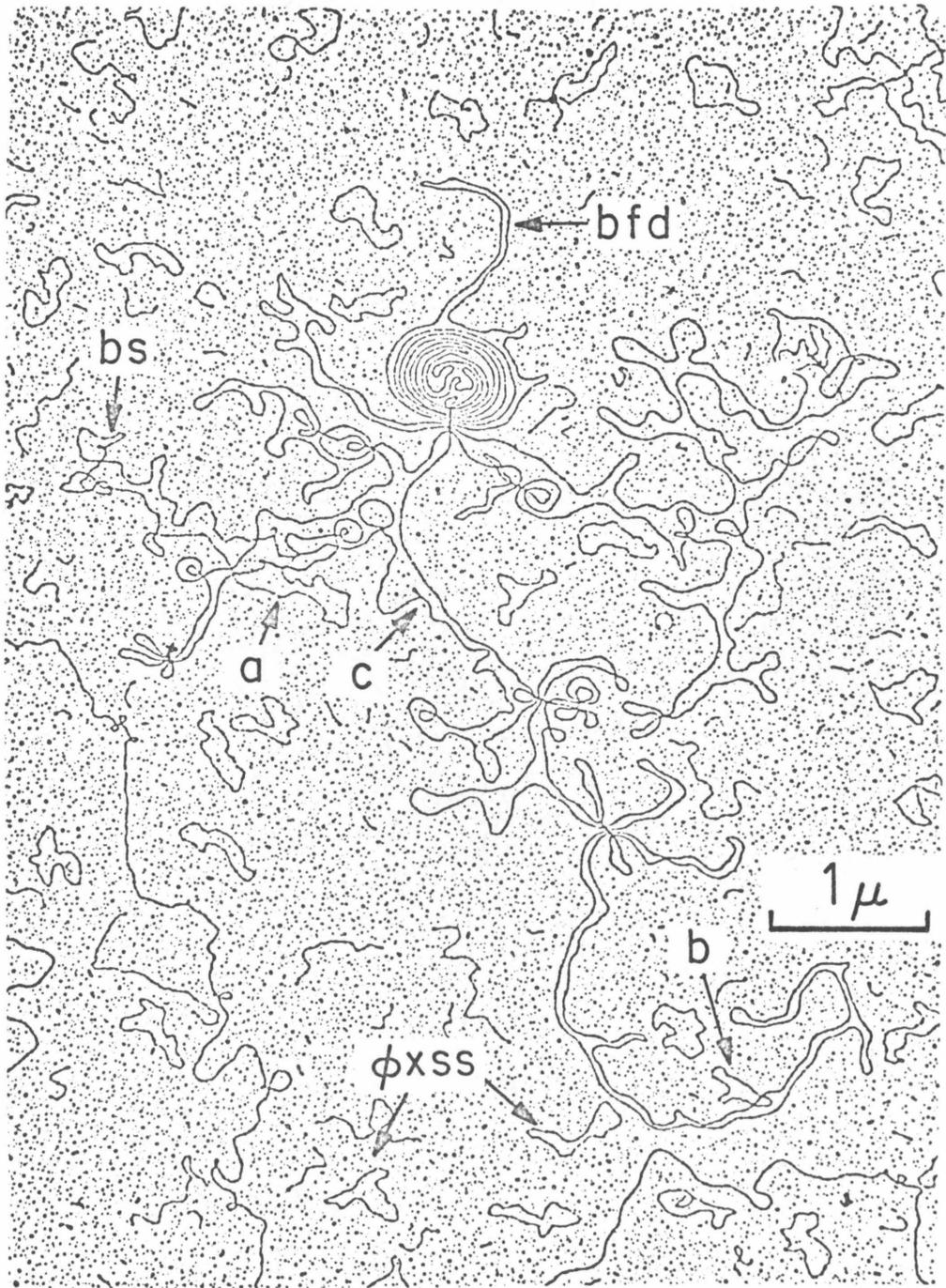


Plate IX

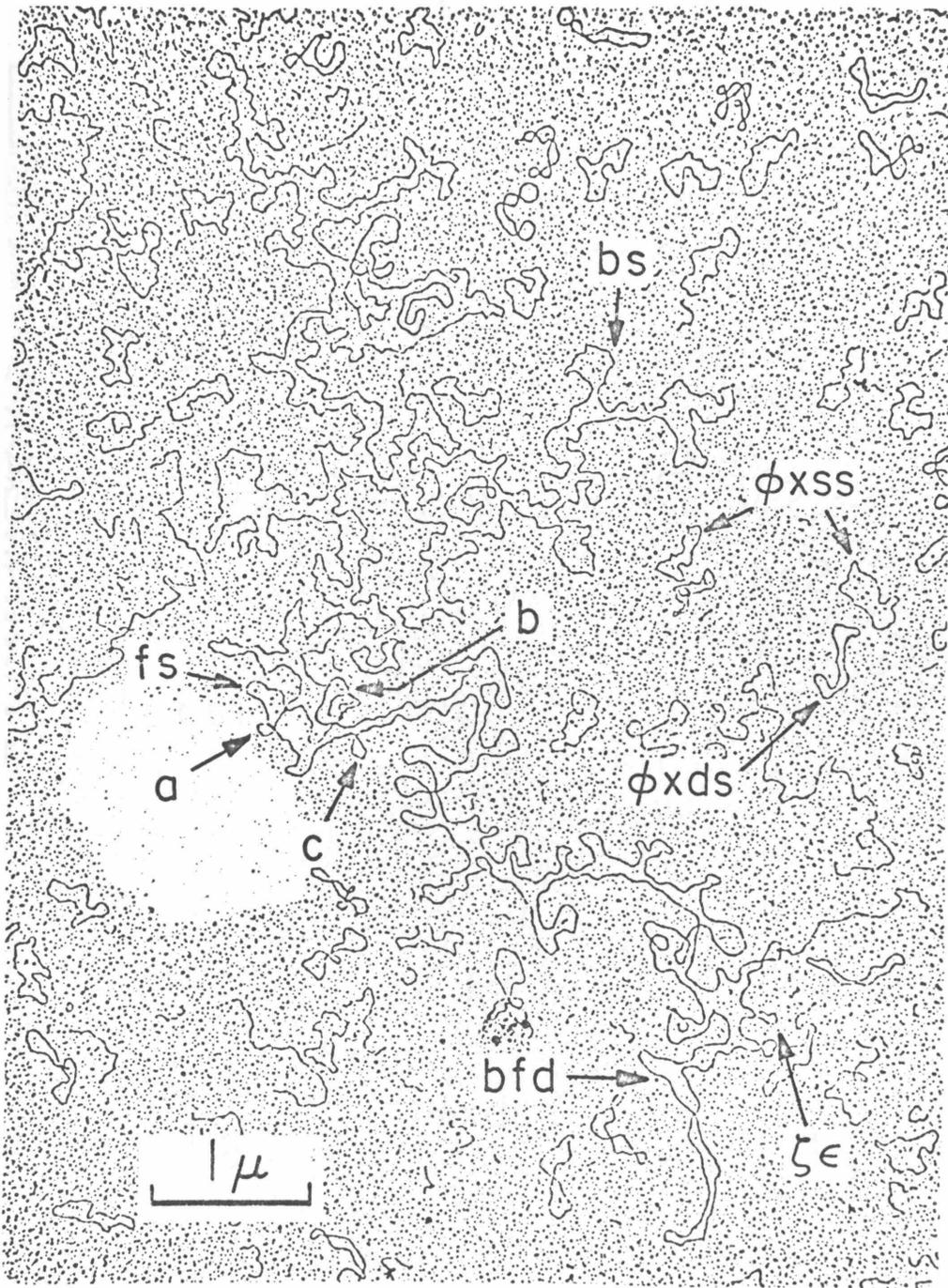


Plate X

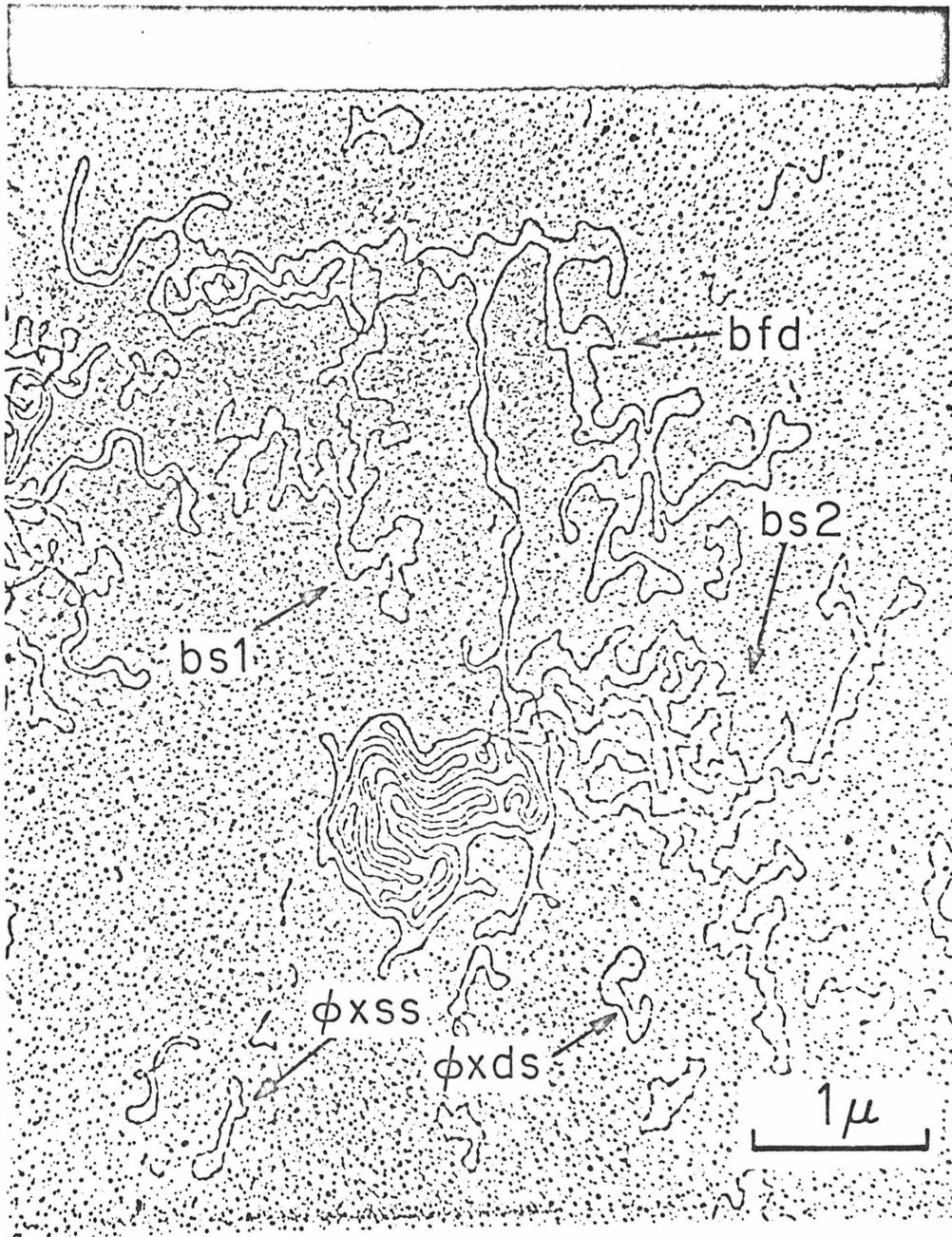


Plate XI

Chapter 3

Electron Microscope Analysis of
Partial Denaturation of F Factor
DNA.

This chapter is a preprint to be
submitted for publication in
Journal of Bacteriology

Electron Microscope Analysis of Partial
Denaturation of F Factor DNA¹

Ming-Ta Hsu²

Department of Chemistry
California Institute of Technology
Pasadena, California USA 91109

1. Contribution No.
2. Present address: Department of Biochemistry,
Stanford University School of Medicine.
Stanford, California 94305 USA

In E. coli, exchange of genetic information between chromosomes is sometimes dependent on the functions of the sex factor F in the donor cell. The F factor is a closed circular DNA molecule with a molecular weight of 63×10^6 daltons (5). About one-third of the sequences in F DNA are essential for fertility; about 17% of F sequences, clustered in one region of the molecule, can be deleted without affecting any known functions (5). The sites on F used for integration to form an Hfr and the sites on the integrated F for excision to form an F-prime factor have all been mapped within this dispensable region for the several F-prime factors studied. (Ref. 5; Ohtsubo, Deonier, Lee, Hu, and Davidson, personal communication for F14, F13, F210 and related episomes).

The base composition of F is about 49% G+C, similar to that of the E. coli chromosome. The distribution of G+C content in F DNA has been shown to be asymmetrical (2). About 10% of the molecule has a G+C content of about 44% while the rest has about 50% G+C. Falkow and Citarella (2) have suggested that the non-uniform base composition may have functional significance. It was proposed that the breaking point, i.e., the point of origin for conjugal transfer of the DNA, lies within the A+T rich region. The high G+C region was shown to contain sequences homo-

logous to E. coli chromosomal DNA. It was suggested (2) that these homologous sequences are used for the integration of F into the bacterial chromosome.

The purpose of the present work is to map the A+T rich regions on the physical map of F (5), and thereby study possible correlations between base composition and biological function.

Materials and Methods

Bacterial strains: The bacterial strains W1485(F), JE3513(F8-33) and ND3(FΔ(0-14.5)) and the structures of the episomes they carry have been described previously (5). KLF253(F152-1) was obtained from B. Low. The structure of F152-1 will be reported elsewhere (E. Ohtsubo and M. T. Hsu, manuscript in preparation).

DNA isolation: The procedure of Sharp et al. (5) was used.

Partial denaturation analysis: Partial denaturation conditions were achieved by raising the formamide concentration, essentially as that described by Davis and Hyman (1).

0.5ug/ml of the DNA to be studied was dialyzed against the following solution: 80% formamide, 5×10^{-2} M NaCl, 5×10^{-3} M Tris, 5×10^{-4} M EDTA, pH 8.5. 50μl of the DNA solution was mixed with 1 to 2 μl of cytochrome c solution (1 mg/ml in 0.1 M Tris, 0.01 M EDTA, pH8.5) and then spread onto a hypophase of 50% formamide and one-tenth concentration of

the electrolyte used in the spreading solution. After waiting for 2 to 3 minutes, the DNA was picked up with a parlodion coated copper grid, stained with uranyl acetate ($5 \times 10^{-5} \text{M}$ in 90% ethanol) and rotary shadowed with platinum. We find that it is important to wait 2 to 3 minutes before picking up DNA after spreading. DNA picked up too soon shows considerable lateral aggregation. This is probably due to aggregation of cytochrome c in high formamide solvents. DNA picked up too late (after 5 minutes) has poorer contrast. The optimal waiting time depends somewhat on the concentration of formamide used. The higher the formamide concentration, the longer one has to wait.

Heteroduplex molecules were formed according to the procedure of Sharp et al. (1) and are spread under the partial denaturation conditions described above.

Micrographs were enlarged and traced on a Nikon projector. The length of DNA in a denatured region was taken as the average of the lengths of the two single stranded branches. In the case of heteroduplexes, only F sequences were measured. The sum of the lengths of denatured and native regions (and the single stranded F DNA in the heteroduplex molecules) was normalized to unity and the length of the denatured region was expressed as a fraction of the length of F. Some of the heteroduplex molecules

studied were incomplete structure, i.e., formed by re-
naturation between incomplete strands. The lengths of
characteristic segments of the molecules are known by
previous heteroduplex analysis (5). Histograms of the fre-
quency with which a particular region became single
stranded under the partial denaturation condition used
were constructed using a length interval of $(1/250)F$. For
a heteroduplex, structural markers of the heteroduplex
were used to align the structures and normalize the lengths.
For the F homoduplex, the partially denatured molecules were
aligned by one of the two largest and most frequently ob-
served denaturation loops.

Results

Fig. 1 shows a histogram of the partially dena-
tured regions of F DNA when spread from 80% formamide,
 $5 \times 10^{-2}M$ NaCl, $5 \times 10^{-3}M$ Tris, $5 \times 10^{-4}M$ EDTA, pH 8.5 at room
temperature ($21^{\circ}C$). An electronmicrograph of a partially
denatured F DNA is shown in Plate 1. Four major peaks
(I, II, III, IV) and two minor peaks (V, VI) can be seen.
An interesting feature is that most of the A+T rich se-
quences are clustered in one general region of F, compri-
sing about 30% of the molecule.

To map the positions of these A+T rich segments on
the physical map of F (5), partial denaturation of hetero-

duplexes between F and several F-prime factors were studied. The F-prime factors used delete different portions of F sequences and therefore are useful for mapping the A+T rich sequences. The results are described below:

The F8-33/F heteroduplex

F8-33 (it was called F₈(N33) in ref. 5) is a transfer defective derivative of F₈ (3,5). F sequences between coordinates 8.5 and 16.3* are deleted in this episome as can be seen in Fig.2a. There are also two insertion loops, c and $\epsilon\zeta$, mapped previously at coordinates 91.0 and 35.2* respectively. The insertion loop c was chosen as a position marker to map the partially denatured regions in the F8-33/F heteroduplex. An electronmicrograph of a partially denatured F8-33/F heteroduplex at the same partial denaturation condition used for F is shown in Plate 2. The histogram of the distribution of partially denatured regions is shown in Fig.3b. It is plotted as a linear map with coordinate 0.0 of F at the left end of the map. Five peaks at coordinates about 4, 19, 25, 30, and 63 were observed. Peak VI in Fig.1 can be readily identified as the peak at coordinate 63 in Fig.3b. By aligning this peak of the histogram of F and that of the F8-33/F heteroduplex, the

* Revised coordinates, R. C. Deonier, private communication

rest of the A+T rich segments of F can be mapped as shown in Fig. 3a. This assignment is confirmed by the analysis of the following heteroduplexes.

The F (0-14.5)/F heteroduplex

F (0-14.5) is a deletion mutant of F with the sequence between 0.0 and 14.5 kb deleted (see Fig. 2b). Therefore if the previous assignment is correct, F (0-14.5)/F should show only four A+T rich regions around coordinates 19, 25, 30, and 63. This is indeed the case as shown in Fig. 3c and Plate 3.

The F152-1/F heteroduplex.

The structure of the F152-1/F heteroduplex is shown in Fig. 2c. The F sequences with coordinates between 0.0 and 2.8* are deleted in F152-1. Since the F sequence around coordinate 2.8 is probably A+T rich as shown in Fig. 3a, point a in Fig. 2c with coordinate 0.0 is chosen as the reference point. The histogram of the partially denatured regions is shown in Fig. 3d. An electronmicrograph of the partially denatured F152-1/F heteroduplex is shown in Plate 4. Five A+T rich regions were observed. They are mapped at coordinates around 10, 19, 25, 30, and 63, respectively. This result is consistent with the previous assignment.

Discussions

The data presented above show that most of the

A+T rich sequences of F DNA are clustered in one part of the molecule, between coordinates 0 and 30. Thus the DNA sequence in this region has several interesting properties: 1) it is A+T rich as shown above, 2) it is used for recombination with the E. coli chromosome for the integration or excision of F (see Table 1), and 3) it is nonhomologous with R(fi⁺) and colV factors. (The region of sequence homology between these plasmids and F extends clockwise from 50F to 94.5/0.0F, where it ends abruptly (4)). It is speculated that this region may be the bacterial DNA acquired by the "original" F factor through repeated genetic interactions.

Falkow and Citarella (2) have suggested that the A+T rich region of F DNA may serve as the natural breaking point in bacterial conjugation. Since different portions of the F sequence between 0 and 30 can be deleted in various episomes (see Table 2) without affecting either the autonomous replication or the fertility functions, the only A+T rich region that could possibly account for such a correlation will be the region mapped at around coordinate 63 (peak VI in Fig.1). It is therefore interesting to note that this A+T region is located in a region between coordinates 43 and 68, where the replication function and the transfer origin of F factor are confined. This is based on the following facts:

1) sequences clockwise from coordinate 68.0 to 94.5/0.0F can be deleted without affecting the autonomous replication or transferability when complemented by R100-1 (5); and 2) sequences between 0.0 and 43.0 can be deleted without affecting either the autonomous replication or transferability (see Table 2). However, a definite correlation between the A+T rich segment VI and the transfer origin and/or the replication origin of F remains to be studied.

Acknowledgments

I am deeply grateful for the generosity of Professor Norman Davidson in providing his research funds (United States Public Health Service, Grant GM 10991). This work was initiated during a discussion with Dr. Donald Robberson. His generosity in providing his unpublished data and his many advices are most appreciated.

TABLE 1

Episome	Integration site	Excision site	Ref.
F100-1	0.0	2.8	5, a
F100-2	0.0	2.8	5, a
F152	0.0	2.8	a
F152-1	0.0	2.8	a
F42	0.0	2.8	5
F8	8.5	16.3	5
F210	8.5	11.5	b
F14	3.0-8.5	-----	c
F13	16.3-17.5	-----	b

- a) E. Ohtsubo and M. T. Hsu, manuscript in preparation.
- b) E. Ohtsubo and S. F. Hu, private communication.
- c) E. Ohtsubo, R. C. Deonier, H. J. Lee and N. Davidson, private communication.

TABLE 2

Episome	F sequence deleted	Ref.
F152	0.0-2.8	a
F100	0.0-2.8	a
F42	0.0-2.8	5
F210	8.5-11.5	b
F8	8.5-16.3	5
F Δ (0-14.5)	0.0-14.5	5
F Δ (33-43)	33-43	c
F8(P6)	0.0-8.5, 68.1-94.5	5
F13-4	16.3-37.2	b

- a) E. Ohtsubo and M. T. Hsu, manuscript in preparation.
 b) E. Ohtsubo and S. F. Hu, private communication.
 c) W. M. Anthony, private communication.

LITERATURE CITED

- 1) Davis, R. W. and R. W. Hyman. 1971. A study in evolution : the DNA base sequence homology between coliphage T7 and T3. *J. Mol. Biol.* 62; 287-301.
- 2) Falkow, S., and R. V. Citarella. 1965. Molecular homology of F-merogenote DNA. *J. Mol. Biol.* 12 : 138-151.
- 3) Ohtsubo, E., Y. Nishimura, and Y. Hirota. 1970. Transfer-defective mutants of sex factors in E. coli. I. defective mutants and complementation analysis. *Genetics* 64: 173-188.
- 4) Sharp, P. A., S. N. Cohen, and N. Davidson. 1973. Electron microscope studies of sequence relations among plasmids of E. coli. II. structure of drug resistance (R) factors and F factor. *J. Mol. Biol.* 75: 235-255.
- 5) Sharp, P. A., M-T. Hsu, E. Ohtsubo, and N. Davidson. 1972. Electron microscope studies of sequence relations among plasmids of E. coli. I. structure of F-prime factors

FIGURE LEGENDS

Fig. 1. A histogram of the partially denatured regions in F factor DNA.

Fig. 2. (a) heteroduplex of F8-33/F. The two small insertions of F8-33 are indicated by c and €. The sequences of F between coordinates 8.5 and 16.3 are missing in F8-33. (b) heteroduplex of FΔ(0-14.5)/F. The F sequences between coordinates 0 and 14.5 deleted in FΔ(0-14.5) is seen as a single-strand insertion-deletion loop. (c) heteroduplex of F152-1/F. In F152-1, the F sequence between 0 and 2.8 is substituted by a bacterial DNA carrying the gal operon.

Fig. 3. (a) The linear form of the histogram shown in Fig. 1. Several relevant coordinates of F DNA are shown in the top of the figure. (b) A histogram of the partially denatured regions in the F8-33/F heteroduplex. The labels can be used to correlate the histogram with the structure seen in Fig. 2a. (c) A histogram of the partially denatured regions in the FΔ(0-14.5)/F heteroduplex. (d) A histogram of the A+T rich regions in F152-1/F. The substitution loop seen in Fig. 2c is labelled by a and b.

PLATE LEGENDS

Plate 1. An electromicrograph of the partially denatured F DNA. A histogram is shown in Fig. 1 and Fig. 3a.

Plate 2. The partially denatured F8-33/F heteroduplex. The substitution loop and the two insertion loops are labelled by a, b, and c and ζ respectively. Five partially denatured regions (I, III, IV, V, VI) can be seen. A histogram is shown in Fig. 3b.

Plate 3. The partially denatured F Δ (0-14.5)/F heteroduplex. The single-strand deletion loop which is the F sequence deleted in F Δ (0-14.5) is indicated by the label d. Four partially denatured regions (III, IV, V, VI) are observed. A histogram is shown in Fig. 3c.

Plate 4. The partially denatured F152-1/F heteroduplex. The molecule shown in this micrograph is formed between incomplete strands of the episomes. Two of the five partially denatured regions, V and VI, are not shown in this micrograph.

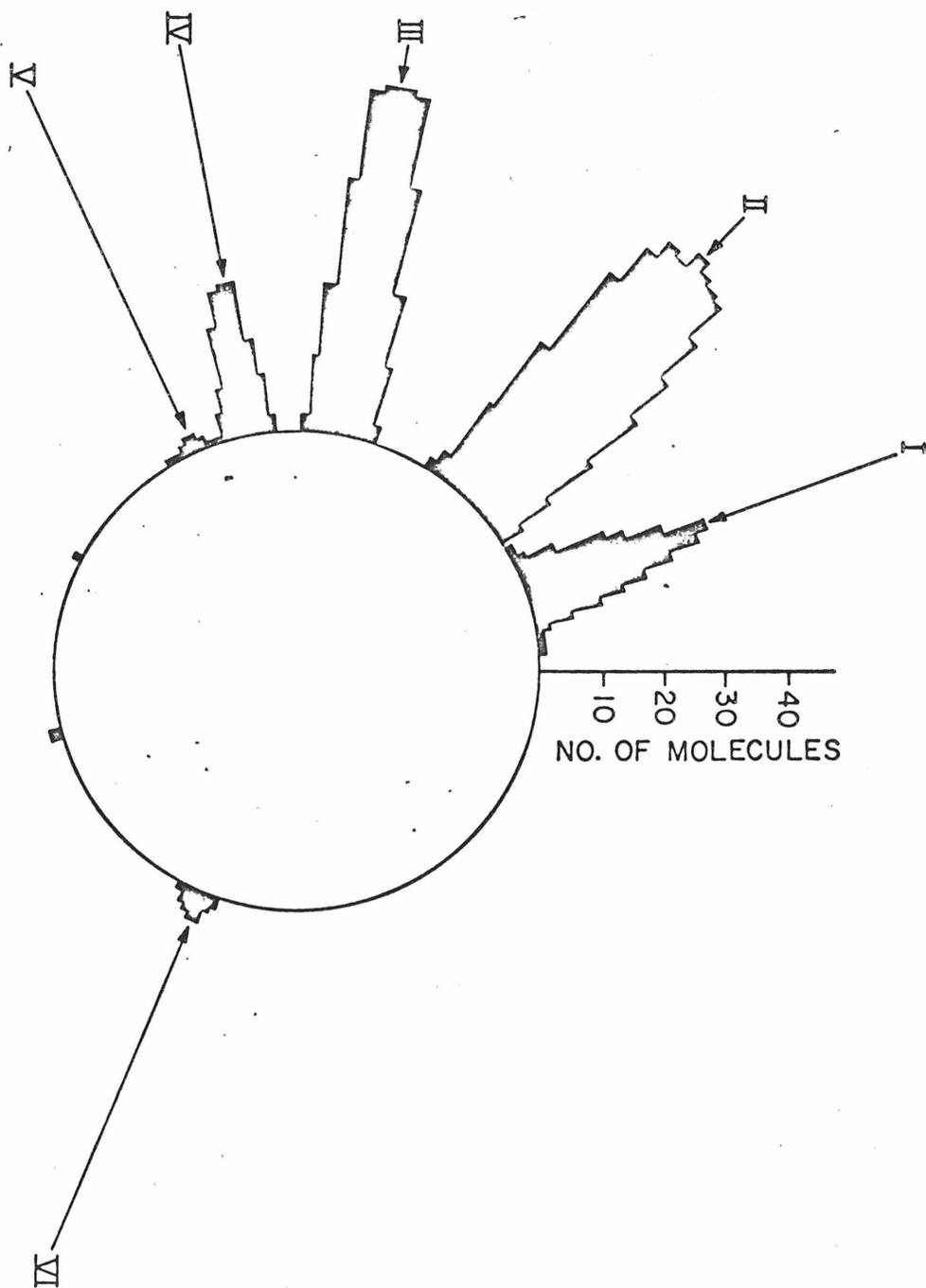


Fig. 1

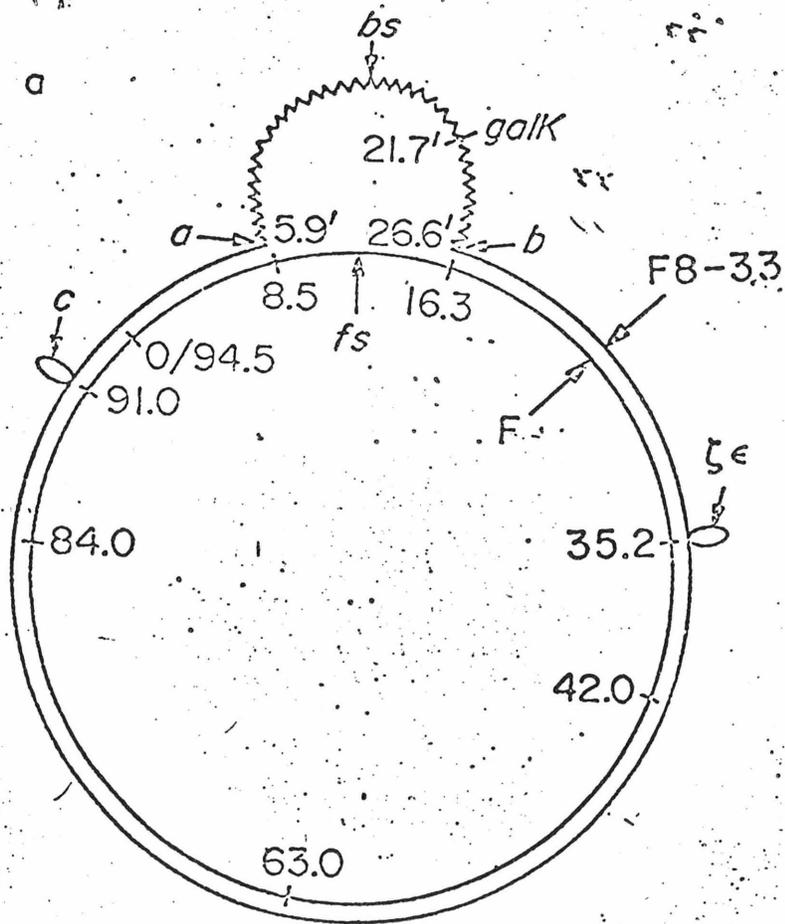


Fig. 2a

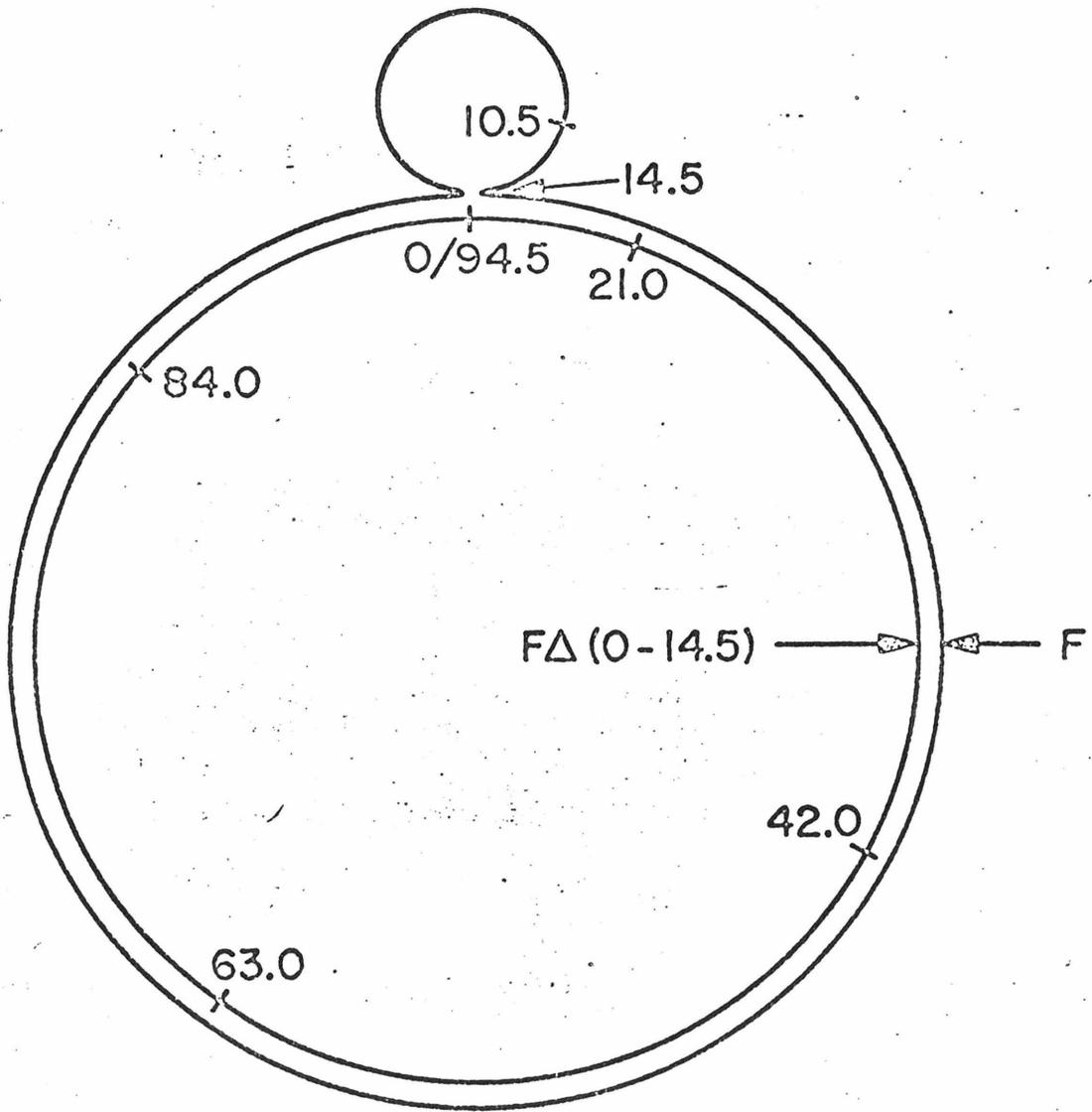


Fig. 2b.

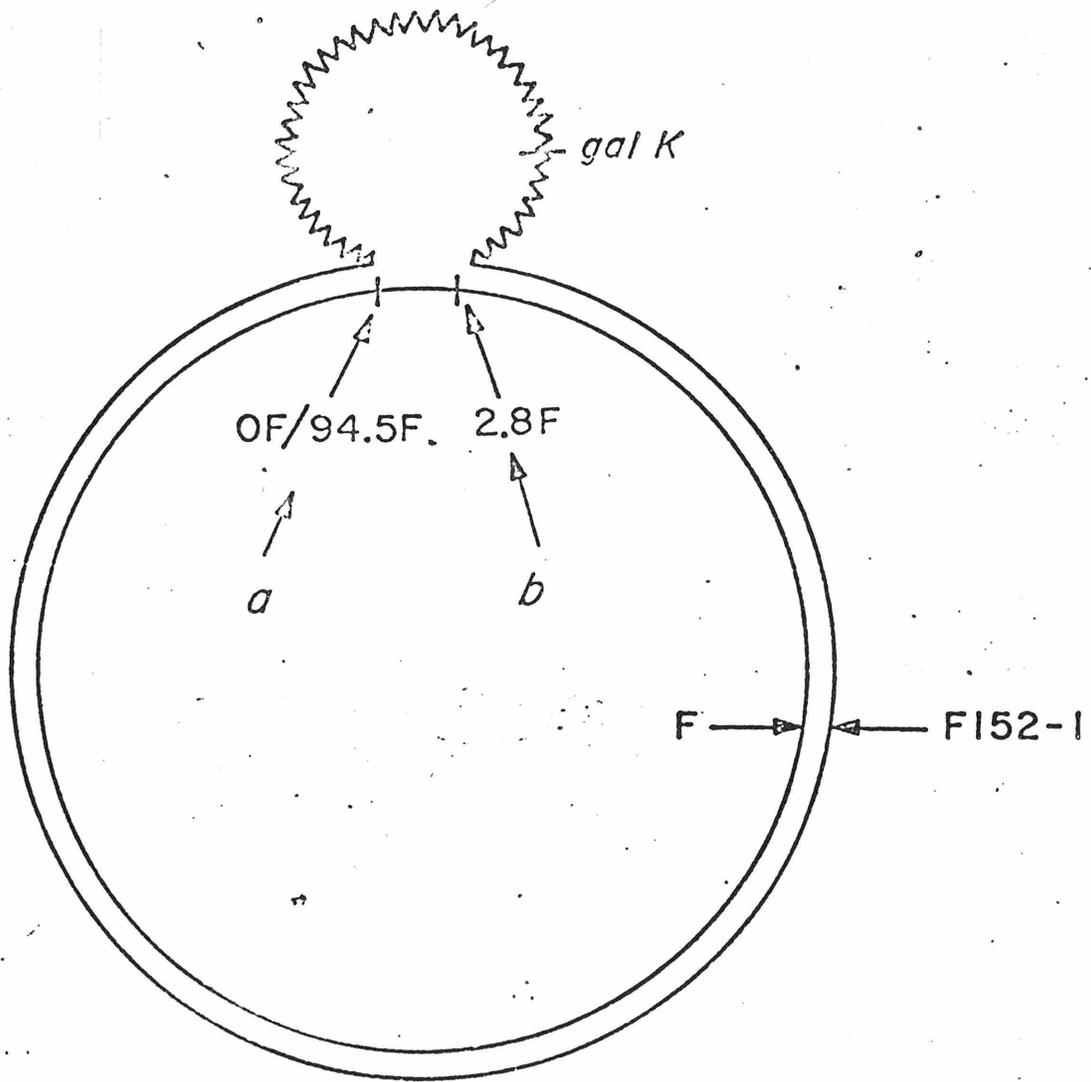


Fig. 2c

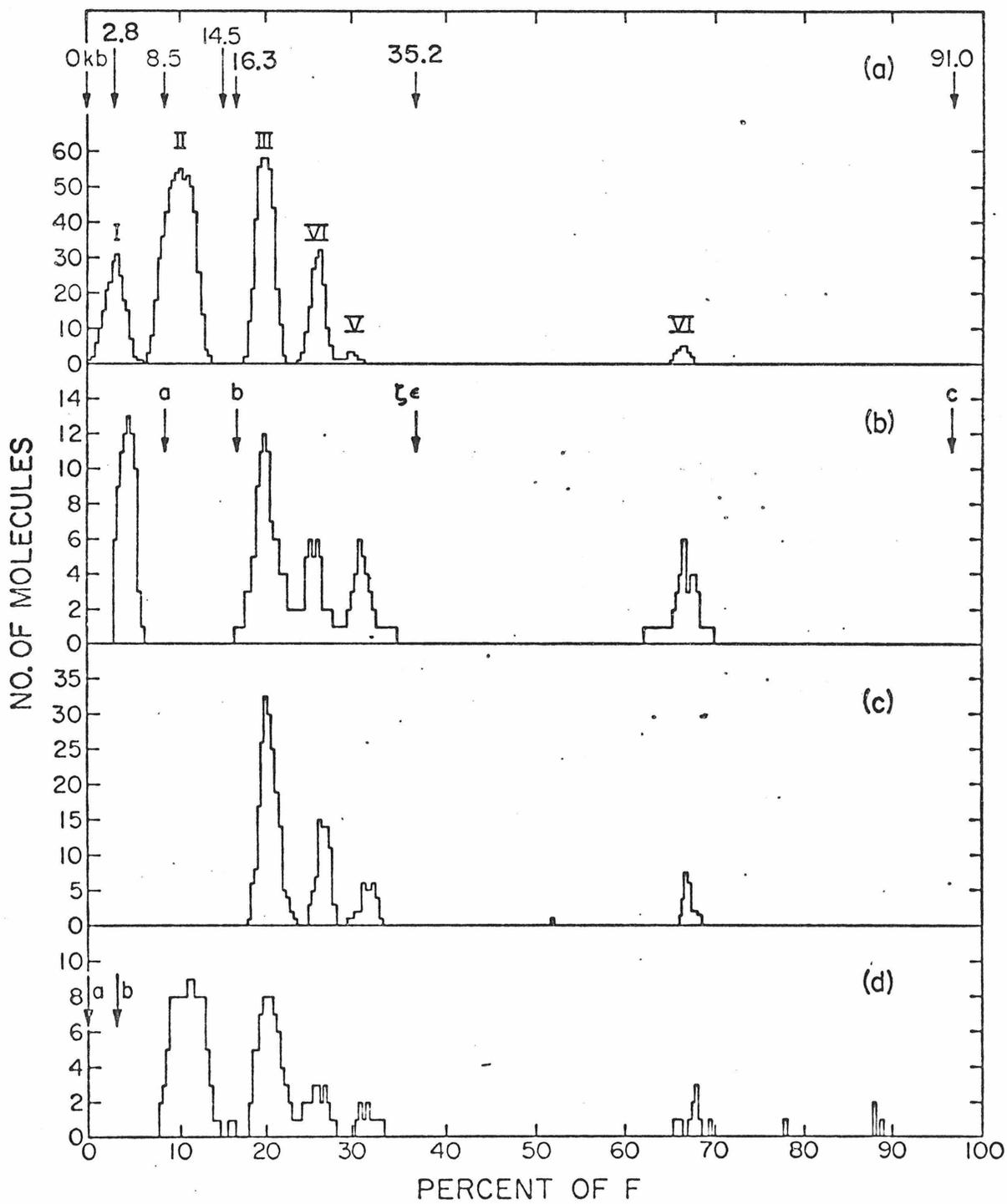


Fig. 3

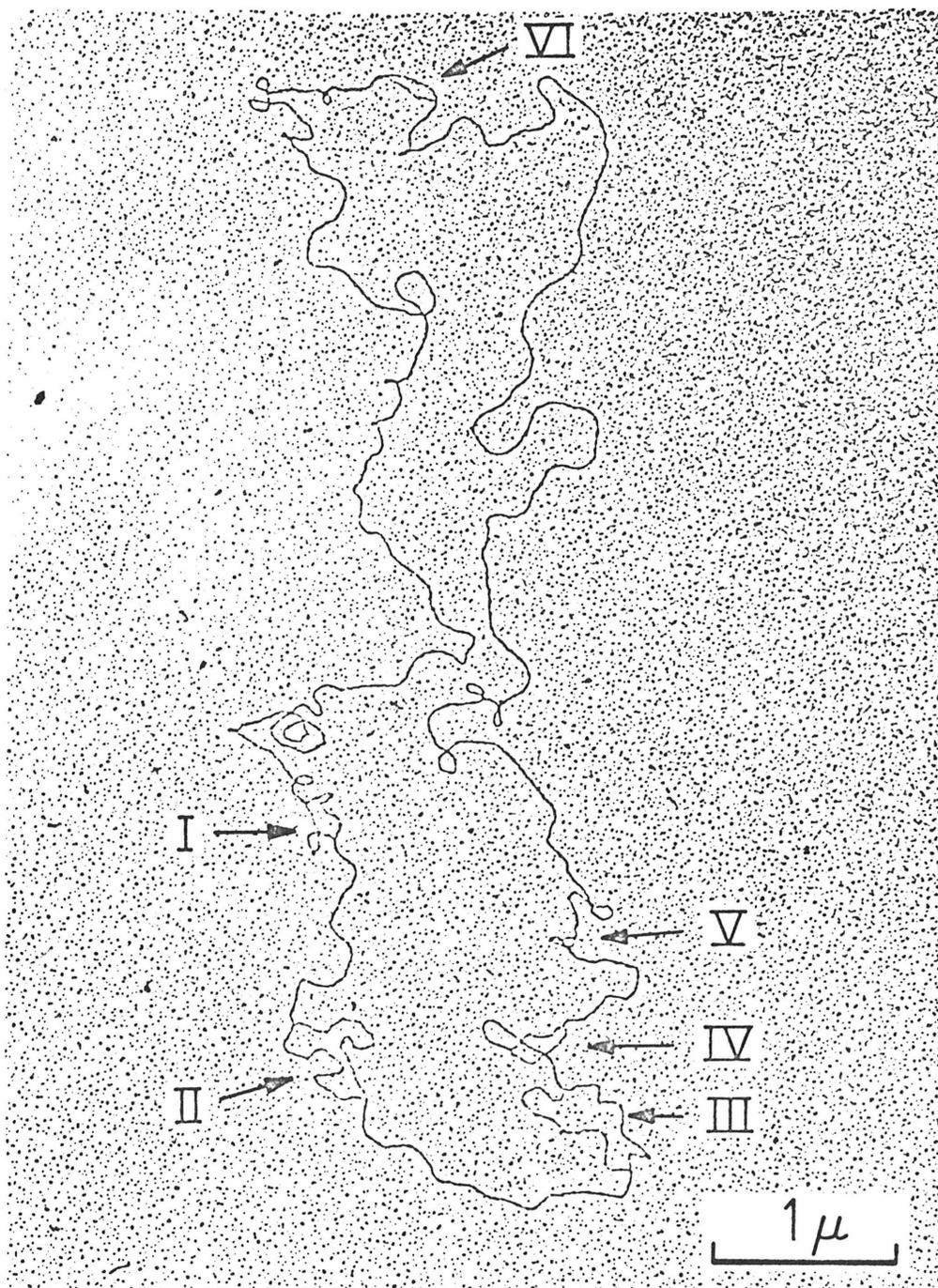


Plate 1

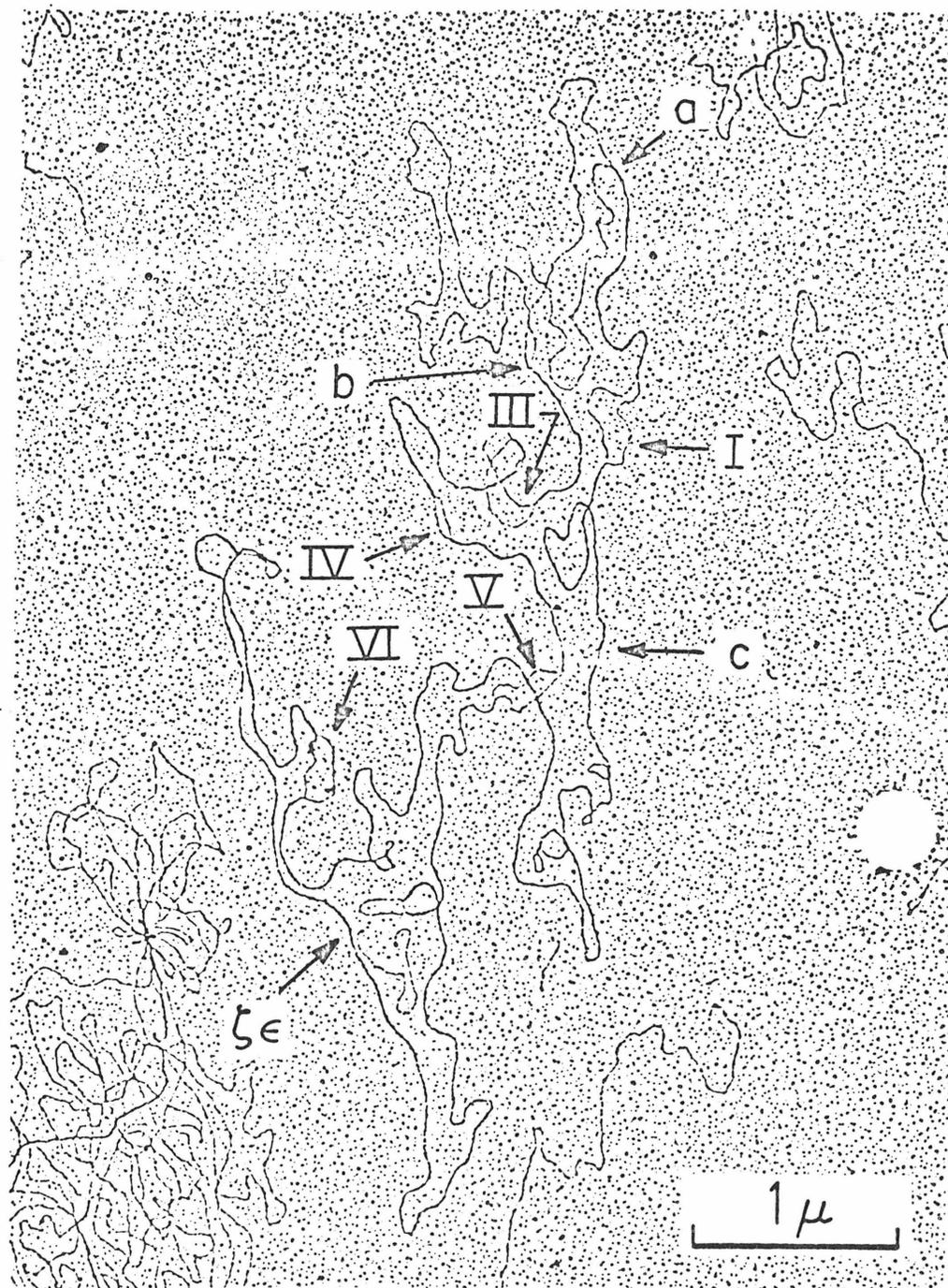


Plate 2

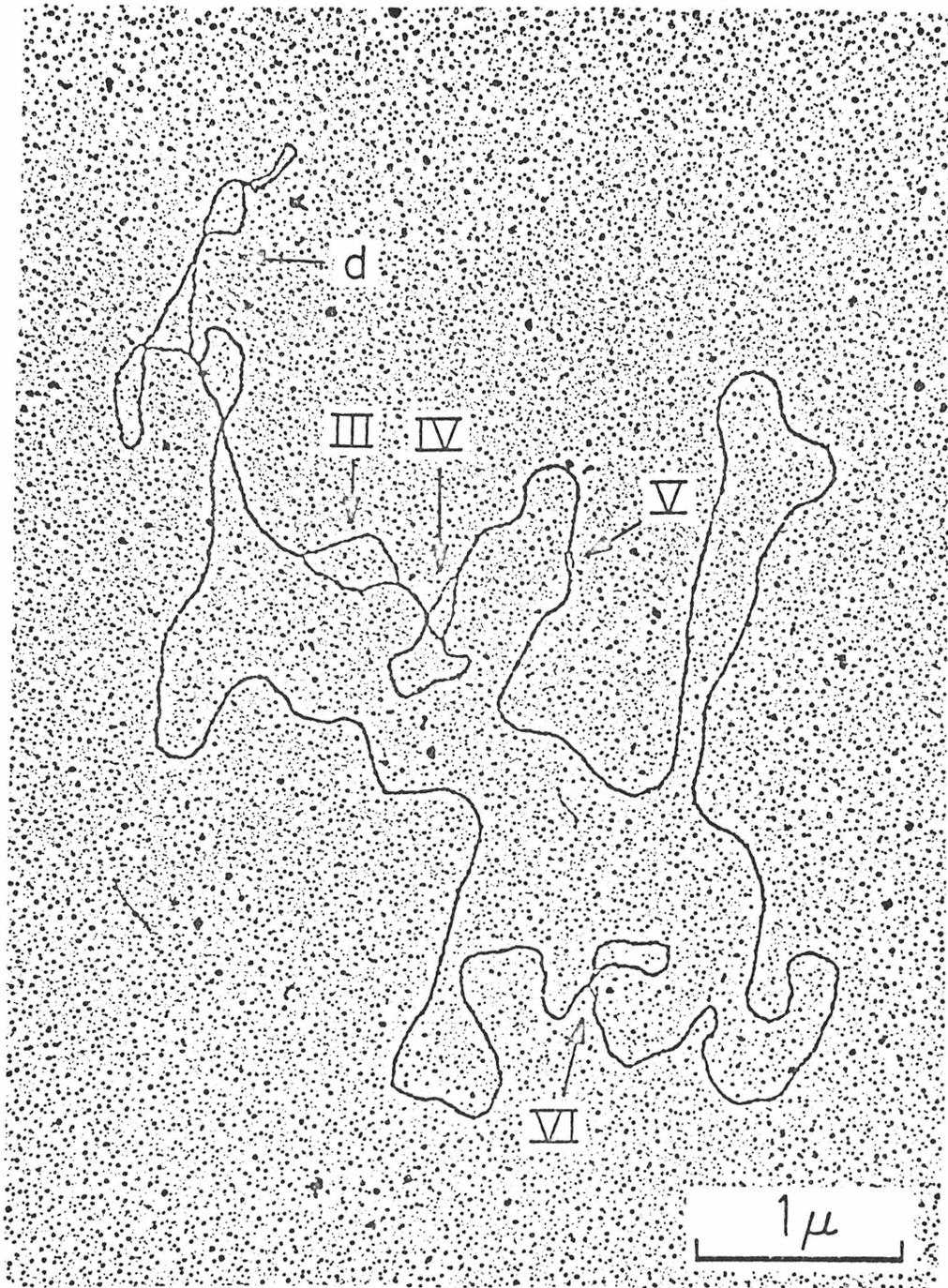


Plate 3

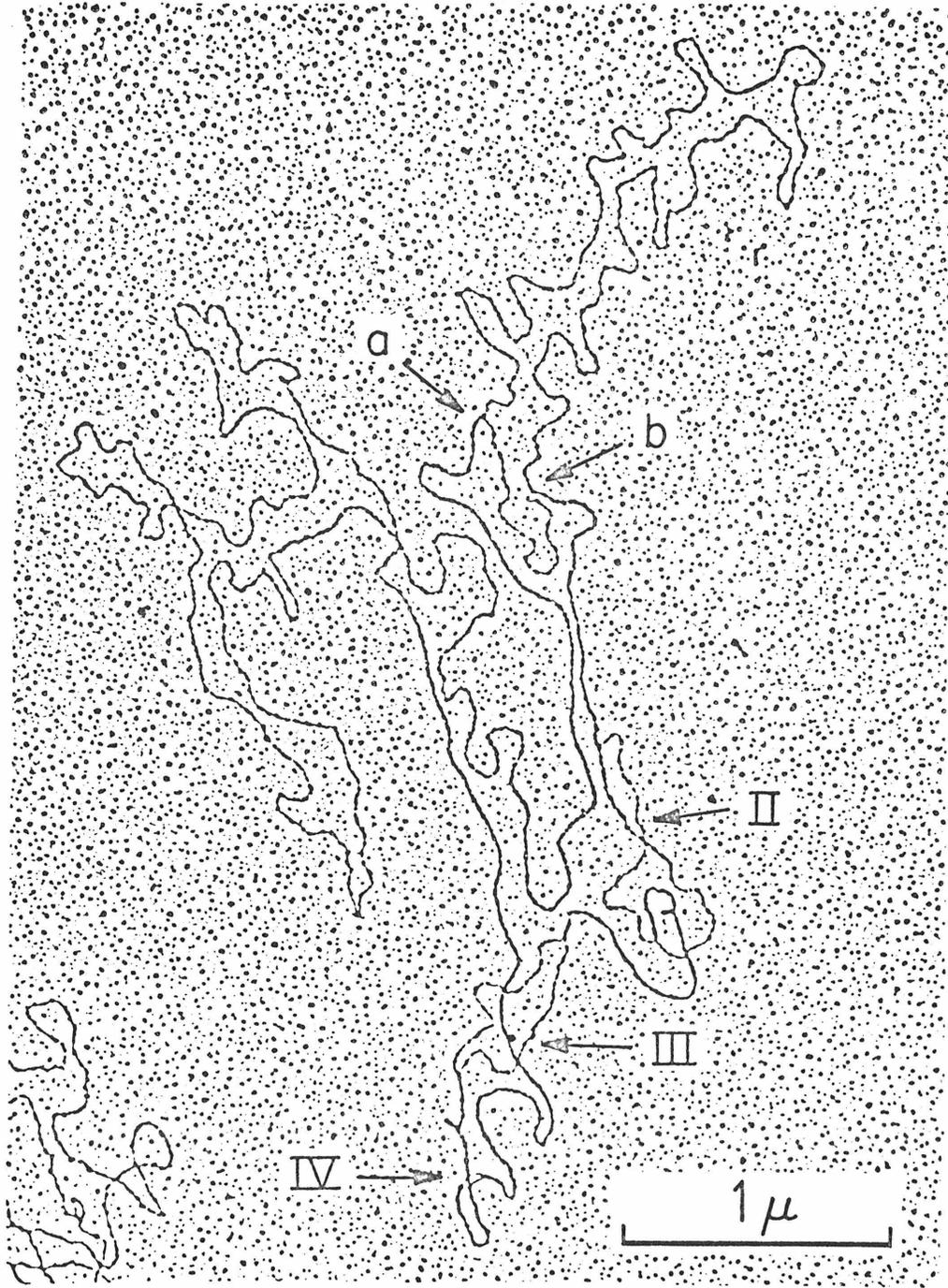


Plate 4

130

Part II

ELECTRON MICROSCOPE STUDIES OF λ
AND MU PROPHAGES

Chapter 1

On the Structure of Prophage λ

LETTERS TO THE EDITOR

Note on the Structure of Prophage λ

Direct physical evidence confirming the Campbell model for the structure of prophage λ was obtained by observing, in the electron microscope, the structure of a heteroduplex between one strand of an episome, F450(λ^{++}), bearing a λ prophage, and the complementary strand of $\lambda b5$ DNA. The facts that the resulting λ hybrid is circular and the concordance of the quantitative contour length measurements with expectation, both confirm the Campbell model. Additional confirmation plus the mapping of the att^{λ} site on F450 are accomplished by constructing the diheteroduplex F450(λ^{++})/F/ $\lambda b2b5c$.

An accompanying paper (Sharp, Hsu, Ohtsubo & Davidson, 1972) describes investigations of the physical structures of the fertility factor, F, and of F-prime factors. In the course of this work, direct evidence for the Campbell model (Campbell, 1962) of the structure of prophage λ was obtained. This result is of interest in itself but is not germane to the main points of our long paper. We accordingly report it here.

In the Campbell model the circularized vegetative phage chromosome inserts by reciprocal recombination into the bacterial chromosome, to insert the phage DNA in a circularly permuted order. This reciprocal recombination occurs between a specific site ($P \cdot P'$ or att^{λ}) close to the middle of the phage DNA and a specific site ($B \cdot B'$ or att^{λ}) on the bacterial chromosome. The substantial evidence supporting this model is reviewed by Gottesman & Weisberg (1971). In particular, physical evidence for linear insertion of λ DNA was provided by Freifelder & Meselson (1970) by showing that there was an increased sensitivity of X-ray nicking of a covalently closed F' molecule after it had been made lysogenic for λ . Rather direct physical evidence for the model as a whole is provided by electron microscope heteroduplex studies of the sequence relationships between λ DNA and the DNA's of λgal and λbio transducing phages (Davis & Parkinson, 1971; Hradecna & Szybalski, 1969). These studies establish the position of the att^{λ} site as $0.573_{\pm 0.01}$ fractional molecular lengths from the left end of λ DNA.

Very direct physical evidence for insertion of the phage genome in a permuted order is provided by our observation of the structure of a heteroduplex, with one strand coming from a bacterial episome carrying an integrated λ prophage and the complementary strand coming from vegetative λ DNA.

The episome, F450, has the bacterial genes *gal*, att^{λ} and *bio* (Freifelder, Folkmanis & Kirschner, 1971; Sharp *et al.*, 1972). λ DNA was integrated into the episome to give F450 (λ^{++}) (Freifelder & Meselson, 1970).

Plate I shows an electron micrograph of a heteroduplex between a strand of F450(λ^{++}) and the complementary strand of $\lambda b5$. The several identifying features are shown in the tracing and in the micrograph. There is a single-stranded circle of F DNA (labeled *fs*). There is a circle, mainly duplex, of λ DNA (λd) leaving the single-stranded circle at att^{λ} . The $i^{\lambda} : b5$ substitution loop occurs, as it must, in the λ duplex with its left terminus, point *m*, at a measured distance of $0.142 (\pm 0.007)$ λ units from att^{λ} ; Davis & Parkinson (1971) and Hradecna & Szybalski (1969) report distances of 0.137 and 0.147, respectively. The measured lengths of the i^{λ} and $b5$ substitution arms are

0.082(± 0.005) and 0.045(± 0.003), and the distance from the right end of the substitution loop around to *att* is 0.78(± 0.01) λ units; all these quantitative data agree with previous measurements. The fact that a circular duplex region of the correct contour length (and with the identifying $i^\lambda : b_5$ loop at the correct distance) forms is direct physical evidence for the insertion of the prophage in a permuted order.

The position of the *att* ^{λ} site on F450 was mapped by constructing a diheteroduplex, F450(λ^{++})/F/ $\lambda b_2 b_5 c$. An electron micrograph is shown in Plate II with an explanatory tracing. The structure of the diheteroduplex is depicted schematically in Figure 1.

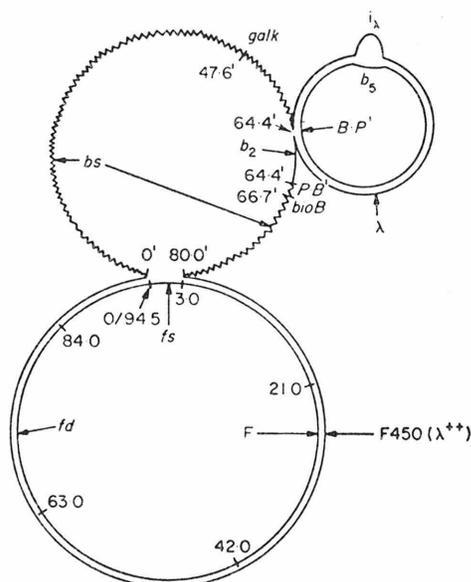


FIG. 1. Schematic representation of the structure of the diheteroduplex F450(λ^{++})/F/ $\lambda b_2 b_5 c$. See also Plate II.

There is a large segment of single-stranded DNA of F450 labeled *bs*; it consists mainly of bacterial DNA but also contains the b_2^+ region of the prophage λ^{++} . There is a circular structure which is duplex except for the identifying $i^\lambda : b_5$ substitution loop due to hybridization of prophage λ^{++} with $\lambda b_2 b_5 c$. This duplex structure joins the single-stranded *bs* at *B·P'*.

The duplex part of the F450/F duplex is labeled *fd*; the piece of F missing in F450 is *fs*. The *B·P'* site of F450 maps at 64.4 kb (thousands of base-pairs) from the counterclockwise junction of bacterial DNA with F DNA. (The evidence for the assignment of the clockwise order of the bacterial DNA sequences is discussed by Sharp *et al.* (1972).) The b_2 region of λ is 5.6 kb long; therefore we assign a point 5.6 kb clockwise from *B·P'* as *P·B'*. (*B·P'* and *P·B'* are assigned the same coordinate of 64.4' since they are a single point on F450 without λ .) The point *P·B'* is measured as 15.6 kb from the clockwise junction of bacterial DNA and F DNA. That is, in F450 (without λ), *att* ^{λ} (i.e., *B·B'*) is 64.4 kilobases from the counterclockwise junction and 15.6 kilobases from the clockwise junction of bacterial DNA and F DNA. The circular structure for the λ part of this heteroduplex and the quantitative agreement of the contour lengths of the several segments again confirm the Campbell model.

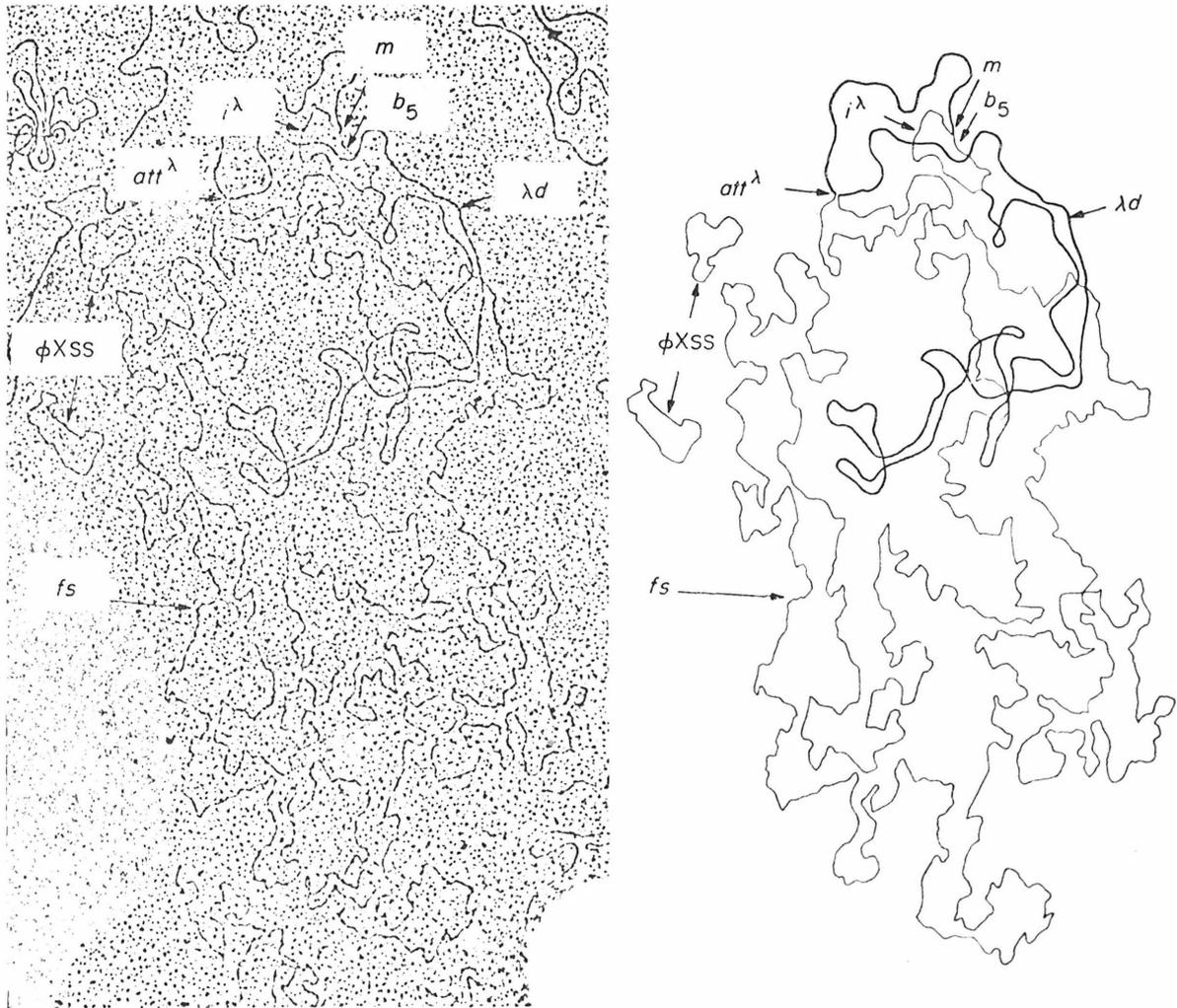


PLATE I. Electron micrograph and explanatory tracing of an F450(λ^{++})/ λb_5 heteroduplex. The structure and the identifying symbols are explained in the text.

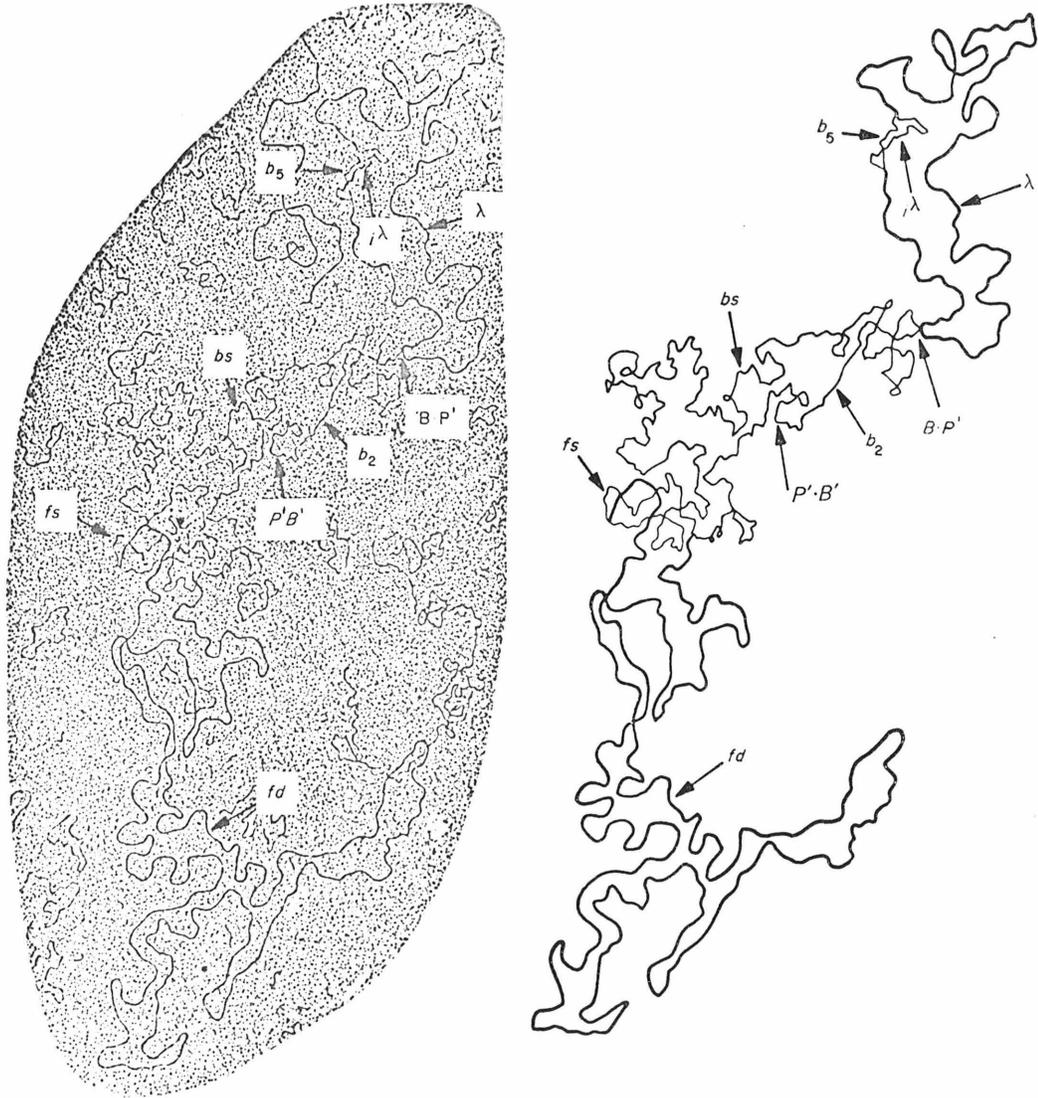


PLATE II. Electron micrograph and explanatory tracing of the diheteroduplex $F450(\lambda^+ +)/F/\lambda 2b5c$. The identifying symbols and the structures are explained in the text.

Department of Chemistry
California Institute of Technology
Pasadena, Calif. 91109, U.S.A.

PHILLIP A. SHARP
MING-TA HSU
NORMAN DAVIDSON

Received 18 February 1972

REFERENCES

- Campbell, A. (1962). *Advanc. Genet.* **11**, 101.
Davis, R. W. & Parkinson, J. S. (1971). *J. Mol. Biol.* **56**, 403.
Freifelder, D., Folkmanis, A. & Kirschner, I. (1971). *J. Bact.* **105**, 722.
Freifelder, D. & Meselson, M. (1970). *Proc. Nat. Acad. Sci., Wash.* **65**, 200.
Gottesman, M. E. & Weisberg, R. A. (1971). In *The Bacteriophage λ* , chapter 6. New York: Cold Spring Harbor Laboratory.
Hradecna, Z. & Szybalski, W. (1969). *Virology*, **38**, 473.
Sharp, P. A., Hsu, M.-T., Ohtsubo, E. & Davidson, N. (1972). *J. Mol. Biol.* **71**, 471.

Chapter 2

Structure of Inserted Bacteriophage
Mu-1 DNA and Physical Mapping of
Bacterial Genes by Mu-1 DNA Insertion

Structure of Inserted Bacteriophage Mu-1 DNA and Physical Mapping of Bacterial Genes by Mu-1 DNA Insertion

(electron microscopy/episomes/*Flac*/*E. coli*)

MING-TA HSU AND NORMAN DAVIDSON

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, Calif. 91109

Contributed by Norman Davidson, July 31, 1972

ABSTRACT It is shown, by electron microscope observation of the structures of heteroduplexes, that Mu-1 DNA inserted into the bacterial episomes *Flac* and $F_2[1]$ is collinear with, rather than a circulation permutation of, the DNA of the mature Mu-1 bacteriophage. Observation of the position of the inserted Mu defines a point within the gene that has been inactivated (the *lacI* gene for *Flac* and a transfer gene in $F_2[1]$ in these particular instances). These examples illustrate a new, general method for physical gene mapping. The episome with Mu DNA inserted into $F_2[1]$ [i.e., $F_2[1](Mu)$], although derived from a single colony, is heterogeneous in that a self-renatured sample shows a nonhomology loop of length 3.0 kb. This nonhomology loop, which has previously been observed in mature Mu-1 DNA, is due to an inversion.

The temperate bacteriophage, Mu-1, induces mutations in its host, *Escherichia coli* K-12. Both genetic and physical evidence support Taylor's hypothesis that a principal mechanism of gene inactivation is the insertion of Mu DNA into the nucleotide sequence of that particular gene (2-6). Mu insertion can take place at a very large number of sites on the *E. coli* genome.

In the present investigation, we have applied the electron microscope heteroduplex method to study the structure of Mu DNA inserted into bacterial F' factors. The results show that inserted (prophage) Mu DNA contains the same sequences in the same relative order as does the DNA of the mature phage; this is to be compared to many other temperate phages, such as the lambdoid phages, where the prophage DNA is circularly permuted relative to the DNA of the mature phage.

The physical position of the inserted Mu on the F' DNA defines the position of the gene inactivated. This is the basis, we believe, for a broadly applicable method for physical gene mapping.

MATERIALS AND METHODS

Bacteria and Bacteriophages. The bacterial strains used are listed in Table 1. M 13 phage was a kind gift from Dr. A. Forsheit.

Abbreviations and notation: kb, kilo bases, a unit of length equal to one thousand bases or base-pairs on single- and double-strand nucleic acids, respectively. We use the notation *Flac*(Mu) with Mu in parentheses in the standard way to indicate that Mu is a prophage borne by the DNA molecule *Flac*. The notation *Flac*(x AT 3155) means that the episome *Flac* is carried in the bacterial strain AT 3155 and/or that the DNA molecule has been extracted therefrom. The notation $F_2[1]$ is an identifying number for one of the F_2 episomes we have studied (ref. 1).

Methods. The methods used to isolate closed-circular, plasmid DNAs and to analyze their sequence relations by the electron microscope heteroduplex method were described (1).

Direct Lysis Procedure. A bacterial culture in 10 ml of L broth (14) was grown to a density of about 5×10^9 cells per ml. The cells were collected by centrifugation at 5000 rpm for 10 min in an SS-34 rotor of a Sorvall centrifuge, washed with 50 mM Tris-5 mM EDTA-50 mM NaCl, (pH 8.5) buffer, and suspended again in a spheroplast-forming solution of 1 mg/ml of lysozyme and 100 mg/ml of sucrose in this buffer. After incubation at 37° for 15 min, an aliquot of the solution was diluted 1:20 into 0.5 M NH_4Ac (pH 7.5), containing 100 μ g/ml of cytochrome *c*. 50 μ l of this solution was spread onto a hypophase of 0.25 M NH_4Ac (pH 7.5), and the protein film was picked up, stained, and shadowed (8).

RESULTS

Mu Inserted into *Flac*. *Flac*(Mu) is an episome carried by strain AT 3156. It contains Mu DNA inserted into the *I* gene of the *lac* operon (3). Length measurements on *Flac*(Mu), on the *Flac* from which it was derived, and on Mu DNA are presented in Table 1. It may be seen from the table that the difference in length between *Flac*(Mu) and *Flac* is 33.4 ± 3.5 kb, which is, within experimental error, the same as the observed length of Mu DNA (37.3 ± 1.1 kb). We thus agree with Martuscelli *et al.* (3) that the two F' factors differ in length by the length of Mu (although our values for the total lengths are somewhat greater). As already pointed out (3), this relation is evidence for the linear insertion of Mu DNA.

In order to study the sequence relations in *Flac*(Mu), the following heteroduplexes and related samples were prepared and studied in the electron microscope: (a) *Flac*(Mu)/*Flac*; (b) Mu DNA self-renatured; (c) *Flac*(Mu)/Mu; (d) *Flac*(x AT 3155)/ $F_2\Delta(33-43)$; (e) *Flac*(Mu)/Mu/ $F_2\Delta(33-43)$.

(a) In the electron microscope, the heteroduplex *Flac*(Mu)/*Flac* consisted of a circular duplex segment of length equal to that of *Flac* plus a single-strand insertion loop of length about equal to that of Mu DNA (Figs. 1 and 2a). This result shows that all of the Mu DNA has been inserted into *Flac* as a single continuous linear structure.

(b) Previous work in this laboratory has shown that the Mu DNA sample used is somewhat heterogeneous (Daniell, E., Abelson, J., Kim, J. P. & Davidson, N., manuscript in preparation; similar observations have been made by H. Delius and E. Bade, personal communication). As depicted

TABLE 1. Properties of bacterial strains and of DNA molecules studied

Strain	Plasmid	Chromosomal genotype	Episomal genotype	Size of episome or phage DNA, $\times 10^{-6}$ daltons	Molecular length, kb	Sources and references
AT 3155	<i>Flac</i>	<i>lac</i> Δ 175 (<i>i</i> ⁻) <i>str</i>	<i>tra</i> ⁺ <i>lacI</i> 78 (<i>i</i> ⁰)	96.6 \pm 1.0	146.4	Taylor, A. L. (3)
AT 3156	<i>Flac</i> (Mu)	<i>lac</i> Δ 175 (<i>i</i> ⁻) <i>str</i>	<i>tra</i> ⁺ <i>lacI</i> 78 <i>lacI</i> 76 (<i>i</i> ⁻) <i>Mu</i> ⁺	118.7 \pm 2.1	179.8	Taylor, A. L. (3)
JE 5303	F ₁ [1]	<i>gal</i> _s <i>trp</i> <i>arg</i> <i>recA</i> <i>str</i>	<i>tra</i> ⁺ <i>gal</i> ⁺ <i>argG</i> ⁺ <i>nadA</i> ⁺	77.8 \pm 2.5	117.9	Ohtsubo, E. (1)
ND 4	F ₁ [1](Mu)	<i>gal</i> _s <i>trp</i> <i>arg</i> <i>recA</i> <i>str</i>	<i>tra</i> ⁻ <i>Mu</i> ⁺ <i>gal</i> ⁺ <i>argG</i> ⁺ <i>nadA</i> ⁺	101.8 \pm 1.2	154.4	Present paper
W 3110(Mu)	—	<i>Mu</i> ⁺	—	24.6 \pm 0.8	37.3	Daniell, E. & Abelson, J. (3, 6)
W 1655	F Δ (33-43)	<i>metB</i> T ₆ ⁸ (λ) ⁻ λ ^r	<i>tra</i> ⁺ ϕ II ⁸	55.9 \pm 3.1	84.3	Broda, P. (7)

in Fig. 2b, a self-renatured Mu DNA sample contains duplex molecules with split ends, SE, and, in some cases, with a substitution loop, G, close to the split ends, as well as perfect duplexes. The significance of this heterogeneity is not fully understood. However, its occurrence is fortunate in that it makes it easy to identify the Mu duplex region in the various *Flac*(Mu)/Mu hybrids described below.

(c) The *Flac*(Mu)/Mu heteroduplex consists of a circle with the contour length of *Flac*(Mu). It is mainly single-strand, but it contains a linear duplex region corresponding to the Mu DNA. Sometimes the duplex region exhibits the same split ends and/or small substitution loop features seen in the self-renatured Mu, thus confirming its identification. The structures are schematically depicted in Fig. 2c, and an electron micrograph is displayed in Fig. 3. This result shows that *Flac*(Mu) contains Mu DNA linearly integrated with its sequences in the same relative order as for the DNA of the mature phage. By contrast, in a heteroduplex of an F' bearing λ prophage DNA with mature λ DNA, a circular duplex region is seen because the sequences of the prophage DNA are a circular permutation of those of the phage DNA (9).

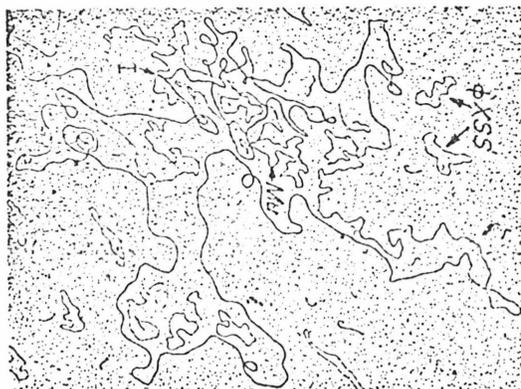


FIG. 1. Electron micrograph of an *Flac*(Mu)/*Flac* heteroduplex. A schematic representation of the structure is shown in Fig. 2a. The point at which the single-strand loop of Mu DNA emerges from the duplex is indicated by I. Due to some single-strand breaks, a segment of the *Flac* duplex circle at the top of the figure is single-stranded. A point on the single strand of Mu DNA is identified. In this and other micrographs, ϕ X-174 DNA molecules of length 5.25 kb are present for length calibration.

(d) The structures of the episomes *Flac* and *Flac*(Mu) were studied by hybridizing each to the episome F Δ (33-43). We found (1) that F(ex W 1485), which is closely related to or identical to the original F of *E. coli* K12, has a length of 94.5 kb, and we have described a coordinate system for specifying the sequences of F. F Δ (33-43) is a deleted F missing the sequence from 32.6 to 42.5 kb (W. M. Anthony, private communication). In heteroduplexes with F or F-prime factors that are not deleted in this region, the readily observable deletion loop provides a reference point for mapping the coordinates of other nonhomology features.

The structure of the heteroduplex *Flac*(ex AT 3155)/F Δ (33-43) is shown schematically in Fig. 2d. It has the following significant structural features: There is a substitution loop in which the longer arm of length 51.7 kb is the bacterial DNA carried by *Flac* but not present in F Δ (33-43), and the shorter arm of length 3 kb is an F sequence missing in *Flac*. The large deletion loop marked B is the F Δ (33-43) deletion and serves as a marker. In addition, there is a small deletion or insertion loop (I-D loop), marked e, of length 1.3 kb at coordinate 22.4 kb on the map of F. All these features have been observed in two other *Flacs* studied in this laboratory, *Flac*(ex JC 1553) and *Flac*(ex DF 3), except that the former doesn't carry loop e (M.-T. Hsu, unpublished results). The three *Flacs* are related, in that they are all derived from a common ancestor *Flac* in 200 P.

(e) A micrograph of an *Flac*(Mu)/F Δ (33-43)/Mu di-heteroduplex is shown in Fig. 4, and schematically depicted in Fig. 2e. The structure has the features expected from the structure of the *Flac*/F Δ (33-43) heteroduplex discussed in the previous section. There is the large marker loop B of F Δ (33-43) at 33 kb and the small I-D loop, e, of *Flac* at 22.4 kb. There is a substitution loop with its shorter arm being the F DNA from 0 to 3 kb that is missing in *Flac*. The longer arm of the substitution loop is longer than the corresponding arm of the *Flac* heteroduplex because of the inserted Mu. The most important point is that the position of the Mu DNA duplex region within this arm is readily observed and measured. Its identification is certain, because in different heteroduplexes all three of the self-renatured Mu structures shown in Fig. 2b were seen; in the micrograph shown in Fig. 4, split ends are seen. The split ends always occur distal to the marker loop B, relative to the junction b. Thus, Mu has been inserted with a defined, fixed polarity. We can see in Fig. 2e that the Mu DNA insertion in the bacterial genes of *Flac*

occurs at a position 45.7 ± 0.8 kb clockwise from the counter-clockwise junction (identified as *a* in the figure) of F DNA with bacterial DNA in *Flac*. This point of insertion of Mu DNA defines a position within the *lacI* gene. The junction *a* defines the point at which F integrated into *E. coli* K12 to form the Υ fr from which *Flac* was derived (1). Thus, a posi-

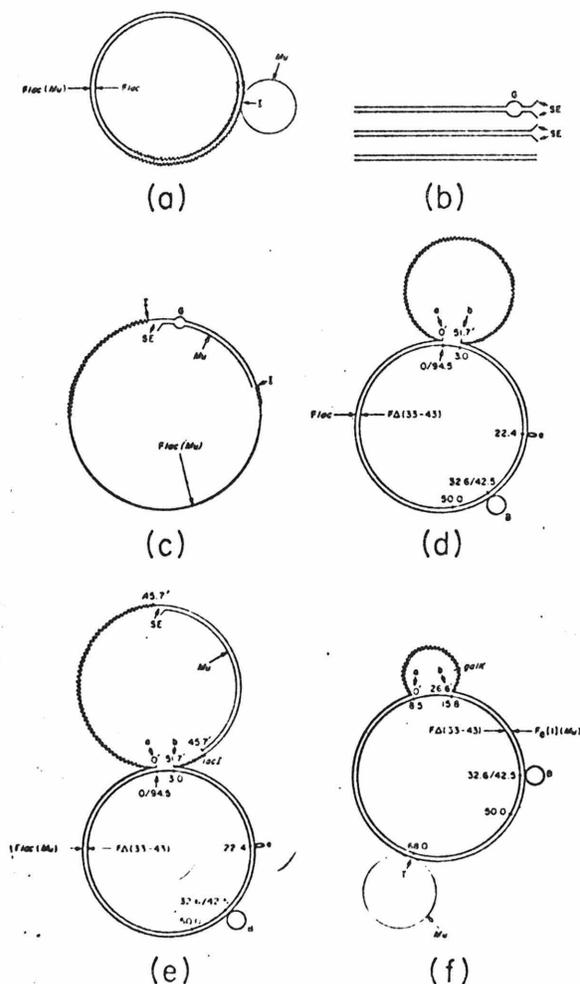


FIG. 2. Schematic representations of heteroduplex structures. (a) The *Flac*(Mu)/*Flac* heteroduplex shown in Fig. 1. *I* is the point at which Mu inserted into *Flac*. (b) Structures seen in self-renatured Mu (Daniell, Abelson, Kim, and Davidson, unpublished); the split ends are indicated by *SE*; the internal non-homology loop of length 3.0 kb by *G*. (c) *Flac*(Mu)/Mu heteroduplex; see also Fig. 3. The split ends, *SE*, and loop, *G*, are indicated. (d) *Flac*/FΔ(33-43) heteroduplex. Straight lines are F sequences, sawtooth lines are bacterial DNA sequences; unprimed and primed numbers are coordinates in kb of F and bacterial DNA sequences, respectively. *a* and *b* are the junctions of F and bacterial DNA. Loop *B* is the 9.9-kb deletion of FΔ(33-43); loop *c* is a characteristic of this *Flac*. (e) The *Flac*(Mu)/FΔ(33-43)/Mu diheteroduplex (see d). The Mu prophage is inserted at coordinate 45.7' kb of the bacterial DNA. (f) FΔ(33-43)/FΔ(33-43) heteroduplex; see also Fig. 5.

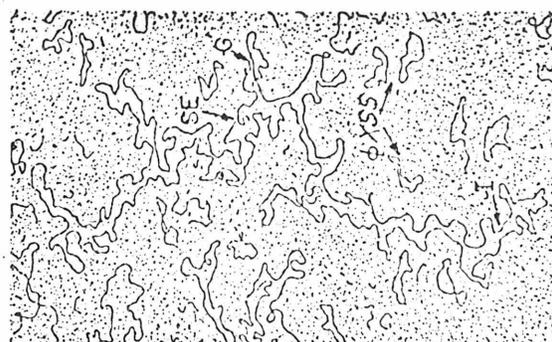


FIG. 3. An *Flac*(Mu)/Mu heteroduplex. Fig. 2c is a schematic structure. The split ends, *SE*, and loop, *G*, are indicated. The other end of the Mu DNA duplex region is indicated by *I* (whereas in Fig. 2c, both ends of the Mu duplex region are labeled *I*).

tion within the *lacI* gene in the segment of the *E. coli* K12 chromosome carried by *Flac* has been physically mapped.

Mu Inserted Within a Transfer Gene of F_Δ[1]. F_Δ[1] harbored in strain JE 5303 is a *gal*-bearing, fertile F' factor (10). Its structure has been described. A culture of strain JE 5303 was treated with Mu; surviving lysogens were selected for *gal*⁺ and M 13 resistance (and thus, a probable loss of fertility). The DNA from such a strain (ND 4) had a contour length equal to that of the parental F_Δ[1] plus one Mu length. The structure of the DNA molecule, as determined from a heteroduplex F_Δ[1](Mu)/FΔ(33-43), is depicted in Fig. 2f. An electron micrograph is shown in Fig. 5. The Mu DNA has been inserted at the coordinate 68 kb on the map of F. This region was identified as one probably containing transfer genes of F (see Fig. 14 of ref. 1), in agreement with the present result that insertion of Mu is correlated with resistance to M 13 phage.

In the F_Δ[1](Mu)/Mu heteroduplex the same Mu structures are seen as in the *Flac*(Mu)/Mu heteroduplex, and in self-renatured Mu. Thus, in F_Δ[1](Mu) as in *Flac*(Mu), Mu is integrated with the same linear order of sequences as in the mature phage.

Intracellular Forms of Mu. We have searched for the kinds of closed-circular molecules observable by direct lysis of a cell bearing Mu inserted in an episome. It is our general experience that when a cell bearing a plasmid is subjected to direct lysis and the DNA is spread for electron microscopy as described in *Methods*, it is possible by careful searching of the grids to recognize twisted circular molecules. The molecules are usually so tightly supercoiled that, in many places, the two duplex filaments are twisted around each other and appear as a single thicker filament. Calibration experiments have shown that the contour lengths of these thicker filament regions times 2.4 plus the lengths of the several duplex regions gives the contour length of the DNA molecule observed as a relaxed circle. By direct lysis of an exponentially growing culture of the strain ND 4, carrying the F_Δ[1](Mu) episome, we observe two size classes of molecules. The major population of molecules had the length expected for F_Δ[1](Mu). The minor population has the length expected for Mu DNA. We propose that this is an intracellular form of Mu.

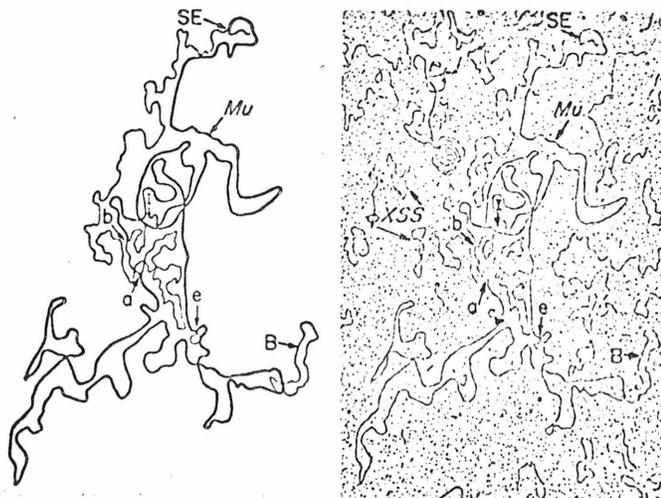


FIG. 4. A *Flac*(Mu)/ $F\Delta(33-43)$ /Mu diheteroduplex, with interpretative tracing. See also Fig. 2c. The point labeled *I* is the same as the point labeled 45.7', *lacI* in Fig. 2c.

Heterogeneity of Mu in $F_8[1](Mu)$. ND 4 bearing the episome $F_8[1](Mu)$ was derived from a single colony. In a self-renatured sample of $F_8[1](Mu)$, perfect circular duplexes were seen. In addition, circular duplexes of the contour length of $F_8[1](Mu)$, but with a single-substitution loop of length equal to 3.0 kb, are observed (Fig. 6). We believe that this nonhomology region is the internal nonhomology loop *G* of Mu (Fig. 2b).

In heteroduplexes $F_8[1](Mu)$ /Mu, molecules of the type shown in Fig. 7 are seen at a low frequency. There is a complete single-strand of $F_8[1](Mu)$ and a complete linear Mu; they are associated by a duplex region of length 3.0 kb at the position of Mu where the nonhomology loop *G* occurs. The interpretation of these findings is that in the prophage in $F_8[1](Mu)$, the sequence *G* of Mu DNA occurs in two states, one the inverse of the other. It should be noted that this inversion occurs even though the bacteria were derived from a single colony and were *recA*⁻. We have further observed

that the same inversion occurs in mature Mu DNA itself; this is the cause of the loop *G* in self-renatured Mu (Fig. 2b).

DISCUSSION

The present physical study shows that Mu DNA inserted into the bacterial chromosome contains the same sequences in the same order as does the DNA of the mature bacteriophage. The prophage is not circularly permuted. We proposed that, after infection, Mu DNA forms a circular molecule. (Our observation of intracellular circular molecules of the correct length, as described above, supports this view.) The circular molecule then inserts into the bacterial chromosome according to the Campbell model (11), but the insertion point is at the junction of the two ends of the mature linear molecule. It is, of course, also possible that the insertion point is a very short distance, perhaps 10-12 nucleotides, away from the ends. Such a short, circularly-permuted sequence in the prophage might be too short to permit cyclization in a hetero-

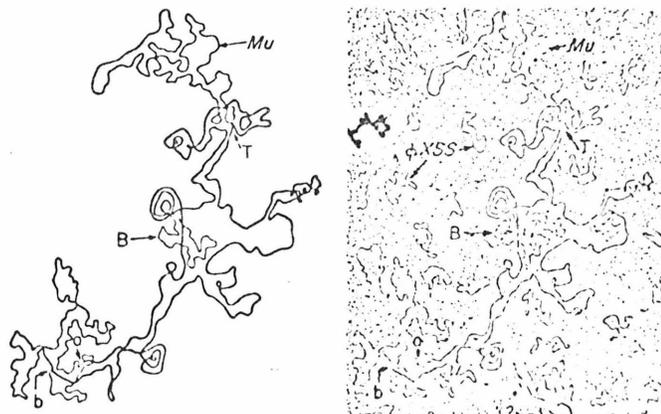


FIG. 5. The $F_8[1](Mu)$ / $F\Delta(33-43)$ heteroduplex with tracing. See also Fig. 2f.

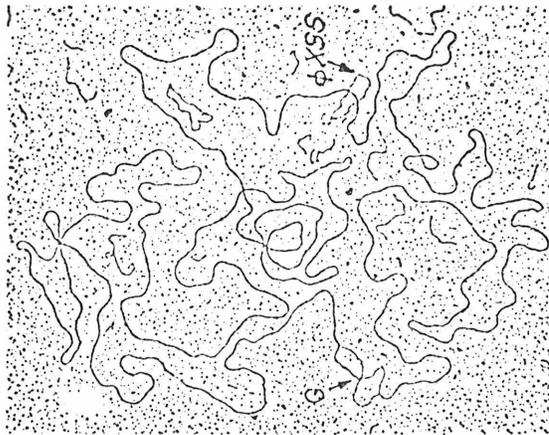


FIG. 6. $F_1[1](\text{Mu})$ self-renatured. The nonhomology loop G is shown.

duplex with the mature phage DNA. [We note that several investigators reported at the Mu Workshop at the Cold Spring Harbor Laboratory in July of 1972, that the genetic maps of Mu phage and prophage are collinear (A. Bukhari, private communication). Thus, the independent genetic and physical studies are in agreement.]

Genetic and/or physical evidence shows that the DNA of the prophages of λ , other lambdoid phages, P 22, etc. are circularly permuted relative to the mature phage DNA molecule (12).

SPO2 and $\phi 105$ are temperate *Bacillus subtilis* bacteriophages. For SPO 2, the genetic map of the prophage is a circular permutation of the mature phage (Truesdell, S., Scibienski, E. & Romig, W. R., private communication); for $\phi 105$, the two genetic maps are collinear (13). L. Chow in this laboratory has provided direct physical evidence that the prophage of SPO 2 is circularly permuted relative to the DNA of the mature phage, by observing a circular structure in prophage-phage heteroduplexes. She interprets the absence of such structures in renatured prophage-phage mixtures for $\phi 105$ to indicate that prophage $\phi 105$ DNA is collinear with that of the mature phage, in agreement with the genetic results.

However, the present observation of a linear duplex structure for Mu DNA prophage-phage heteroduplexes is the first direct physical evidence for a case of lysogeny by insertion without circular permutation.

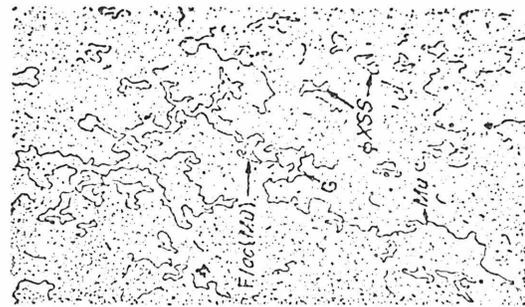


FIG. 7. A linear strand of Mu and a circular strand (actually broken at one point) of $F_1[1](\text{Mu})$ have associated via the duplex segment G only. This and related molecules show that nonhomology loop G in $F_1[1](\text{Mu})$ and in Mu is due to an inversion.

We are grateful to Ellen Daniell and Prof. John Abelson of the University of California at San Diego and to Prof. A. L. Taylor of the University of Colorado Medical School for educating us and stimulating our interest in the properties of Mu-1 bacteriophage DNA. The sample of Mu-1 DNA used in our studies was a generous gift from John Abelson and Ellen Daniell. The first observations in our laboratory of the heterogeneous structure of Mu DNA, as observed in self-renatured samples, were made by J. P. Kim, E. Daniell, and J. Abelson. We appreciate the guidance and advice of E. Ohtsubo throughout this work. This research has been supported by Grant GM 10991 from the NIH. This is Contribution No. 4502 from this Institute.

1. Sharp, P. A., Hsu, M.-T., Ohtsubo, E. & Davidson, N. (1972) *J. Mol. Biol.*, in press.
2. Taylor, A. L. (1963) *Proc. Nat. Acad. Sci. USA* 50, 1043-1051.
3. Martuscelli, J., Taylor, A. L., Cummings, D. J., Chaplin, V. A., DeLong, S. S. & Tanedo, L. (1971) *J. Virol.* 8, 551-563.
4. Boram, W. & Abelson, J. (1971) *J. Mol. Biol.* 62, 171-178.
5. Bukhari, A. I. & Zipser, D. (1972) *Nature New Biol.* 236, 240-243.
6. Daniell, E., Roberts, R. & Abelson, J. (1972) *J. Mol. Biol.* 69, 1-8.
7. Broda, P. (1967) *Genet. Res.* 9, 35-47.
8. Davis, R. W., Simons, N. & Davidson, N. (1971) *Methods Enzymol.* 21D, 413-428.
9. Sharp, P. A., Hsu, M.-T. & Davidson, N. (1972) *J. Mol. Biol.*, in press.
10. Ohtsubo, E. (1970) *Genetics* 64, 189-197.
11. Campbell, A. (1962) *Advan. Genet.* 11, 101-145.
12. Scaife, J. (1967) *Annu. Rev. Microbiol.* 21, 601-638.
13. Armentrout, R. W. & Rutberg, L. (1970) *J. Virol.* 6, 760-767.
14. Meynell, G. G. & Meynell, E. (1970) *Theory and Practice in Experimental Bacteriology* (Cambridge University Press, New York), 2nd ed.

Chapter 3

Electron Microscope Hetero-
duplex Study of the Hetero-
geneity of Mu Phage and pro-
phage DNA

This chapter is a preprint to
be submitted for publication in

Virology

Electron Microscope Heteroduplex Study of the
Heterogeneity of Mu-Phage and Prophage DNA^{1,2}

MING-TA HSU³ and NORMAN DAVIDSON⁴

Department of Chemistry
California Institute of Technology
Pasadena, California 91109, U.S.A.

SUMMARY

The nature of the G-loop heterogeneity and of the split end heterogeneity in Mu phage DNA have been studied. The structures of the molecules seen in self-renatured Mu phage DNA samples show that there are two classes of native molecules in which the G sequence occurs either in a direct or in an inverted order. By studying the structure of a single strand of Mu DNA when spread under weakly denaturing conditions, it is shown that the G region is flanked by very short (50 nucleotides or less) inverted repeat sequence. The G loop inversion is attributed to a phage

1. Contribution No.

2. Supported by Grant GM 10991 from the United States Public Health Service.

3. Present address: Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, U.S.A.

4. To whom reprint requests should be addressed.

enzyme directed reciprocal recombination between these inverted repeat sequences. A second slightly larger inverted repeat sequence occurs within the G loop, but this sequence is not active in reciprocal recombination. By studying suitable diheteroduplexes of F-prime factors containing an inverted Mu prophage, with a reference F factor and with Mu phage DNA, it is shown that the highly heterogeneous split ends of Mu phage are lost when Mu becomes a prophage.

INTRODUCTION

The structures of the molecules observed after denaturation and renaturation of phage Mu DNA (Fig.1a and b) show that there are two kinds of sequence heterogeneity in this DNA (Daniell, Abelson, Kim, and Davidson, 1973). All, or almost all, of the renatured molecules have a nonhomologous sequence at one end (the split ends,SE); about 50% of the molecules show a symmetrical loop (the G loop) near SE.

In a previous communication (Hsu and Davidson, 1972), we showed that the G-loop heterogeneity is also present in Mu-prophage DNA and that it is due to an inversion.

In the present short communication, we report evidence that the G loop in the phage DNA is also due to an inversion. Further pertinent information about the sequence in and bordering the G loop is reported. We also show that the

split-end sequences are lost when Mu DNA inserts and becomes a prophage.

MATERIALS AND METHODS

Bacterial and phage strains. Strains bearing the episomes Flac, Flac::Mu, and F8(68F::Mu) (which we called Flac, Flac(Mu) and F₈[1] (Mu) in Hsu and Davidson, 1972) are described in Table I of our previous paper (Hsu and Davidson, 1972). In particular, it should be recalled that strain ND4 bears an episome F8(68F::Mu) in which Mu is inserted into one of the tra genes of F8 (This F-prime was previously called simply F8[1]). ND5 is another derivative of JE5303, selected as a Mu lysogen. It contains Mu inserted into the F part of the F8 episome at F coordinate 45 kb; there is no known mutation caused by this insertion. This episome is denoted as F8(45F::Mu).

Electron microscopy. Our methods for preparing DNA samples and for the electron microscope heteroduplex analysis have been described previously (Sharp et al., 1972).

RESULTS AND DISCUSSION

The nature of the G loop. By careful examination of samples of denatured and renatured Mu DNA the heteroduplex of Fig. 1(c) is seen at a low frequency. A micrograph is

shown in Plate I. A model for the sequence organization of native Mu phage DNA which explains this structure and the presence of the G loop in Fig. 1(b) is shown in Fig. 1(d). In the figures, a sequence and its complement are denoted by unprimed and primed letters, respectively. Mu-DNA molecules of type 1 differ from Mu-DNA molecules of type 2 in that the G sequence is inverted in one relative to the other. The heteroduplex of Fig. 1(b) contains, for example, a W_2 strand and a C_1 strand (or a C_2 strand and a W_1 strand). The heteroduplex of Fig. 1(c) consists of a W_1 strand mated with a W_2 strand. Only the G sequences are complementary. The observed lengths of the G duplex region and of the several single-strand regions are quantitatively consistent with this interpretation. The probability of formation of a $W_1:W_2$ heteroduplex molecule is much less than that of $W_2:C_1$ molecule because of the much shorter length of the complementary sequences.

As discussed below, a plausible model for the generation of the inversion is the presence of an inverted repeat sequence at the two borders of the G region. In order to look for such a sequence, Mu DNA was denatured, reneutralized, and immediately spread, without allowing time for renaturation of complementary strands. The single strands were spread under less denaturing conditions (30% formamide;

0.1 M Tris, 0.01 M EDTA, pH 8.5 onto 5% formamide, 0.01 M Tris, 0.001 M EDTA) than usually used for extending single strands. The structure shown in Fig. 2 and Plate II was observed quite frequently. It has the following characteristics. In a single strand of Mu DNA there is a loop of fixed length occurring at a fixed position from the end of the molecule. The length and position of the loop is consistent with the hypothesis that it consists of the G loop sequences. Within the loop, there is a second duplex region and two single-strand regions. Figure 2 also presents a model to interpret this structure. The G loop is bordered by the inverted repeat sequences g_1 and g_1' . These mate with each other under weakly denaturing conditions, thus causing the formation of the large loop. The g_1 sequence must be quite short, perhaps less than 50 nucleotides, since the duplex $g_1:g_1'$ appears as a point in the micrographs. In addition to the unrepeated sequences, g_2 , g_4 , and g_5 the G loop also contains an additional inverted repeat sequence g_3 and g_3' disposed as indicated in the figure. Its length is about 300 nucleotides. The total length of the sequences g_2 , g_3 , g_4 , g_3' and g_5 is observed to be equal to the size of the G loop in the heteroduplexes of Fig. 1. The $g_3:g_3'$ inverted repeat structure is also seen in the G loop of the structure of the type Fig. 1(b) when spread under suitable

formamide condition (see Plate IIIa, b). This positively shows that the inverted repeat sequences are a part of the G loop. The observation that the structure is present in both arms of the G loop confirms our previous interpretation that the G loop heterogeneity is due to a sequence inversion.

We propose that the occurrence of the G loop inversion in the DNA derived from a single colony of a Mu lysogen arises from reciprocal recombination between the inverted repeat sequences g_1 and g_1' flanking the two sides of the G loop. This is illustrated in Fig. 3. Synapsis of the two g_1 sequences and subsequent recombination will give rise to the inversion. The occurrence of the inversion is known to be independent of recA, B, and C functions (Daniell, Boram, and Abelson, 1973); therefore the recombination is probably not caused by phage coded recombination enzymes.

Under weakly denaturing conditions, structures of self-renatured Mu phage DNA and Mu prophage DNA as shown in Fig. 4a, b and Plate IIIc, d were also observed. They were formed by the renaturation of the complementary g_3 sequences in the two arms of the G loop. Structures shown in Fig. 4c, d, formed from the structures shown in Fig. 4a by the inversion of g_4 sequence, were not seen. Therefore reciprocal recombination between the inverted repeat sequences g_3 and g_3' has never been observed. A perfectly matched duplex

sequence of length 300 nucleotides and of average 50% G+C base composition should remain duplex when spread from 50% formamide, 0.1 M Tris, 0.01 M EDTA. But the $g_3:g_3'$ duplex is not seen under these circumstances. Thus, either there is partial mismatch between these complementary sequences or they are very A+T rich. The absence of recombination at this inverted repeat is either due to partial sequence mismatch or due to sequence specificity of the phage recombination system.

Split end sequences in Mu prophage. It should be recalled that the present evidence is that the split end sequences, SE, of Mu phage DNA are very heterogeneous and differ from molecule to molecule (Daniell, Boram, and Abelson, 1973; Delius and Bade, private communication). One plausible hypothesis is that the split ends are heterogeneous pieces of bacterial DNA picked up by Mu during its life cycle. We now wish to ask whether these split end sequences are present in Mu prophage. Hybridization of Mu prophage with Mu phage DNA shows that the former contains the G loop inversion (Hsu and Davidson, 1972). When an episome containing Mu prophage is denatured and self-renatured and spread from 50% formamide (so that the $g_3:g_3'$ inverted repeat duplex is unstable), circular molecules showing the G loop and perfect duplex circular molecules are seen. The former structure is shown in Fig. 5. If the SE nonhomology region was present in Mu prophage, a circular duplex with two loops, as depicted in

Fig. 5, should also be seen. But two loop structures are not observed. This means either that all the prophage DNA isolated from a purified lysogen is homogeneous with respect to the split end sequence or the split end sequence is simply not present in the prophage. The evidence presented below supports the latter interpretation.

Three independently isolated F-prime factors carrying Mu prophage were studied: (1) Flac::Mu in strain AT3156; (2) F8(68F::Mu) in ND4 (both as described by Hsu and Davidson, 1972); and (3) the new F8(45F::Mu) episome, carried in strain ND5. In this episome the prophage Mu DNA is inserted in the F sequence at F coordinate 45 kb as depicted in Plate IV

The following diheteroduplexes were constructed for analyzing the structure of the prophage in these episomes: Flac::Mu/Flac/Mu; F8(Mu)/FΔ(33-43)/Mu (F8(Mu) denotes either of the F8 episomes, F8(68F::Mu) or F8(45F::Mu), bearing Mu as a prophage). Consider the latter case. It should be recalled that FΔ(33-43) is a mutant of F, deleted in the 33 to 43 kb sequences. It is useful as a reference molecule in heteroduplex studies of F-primes. In both F8(Mu) episomes, the Mu DNA is inserted in F sequences of F8, and at a point which is not within the 33 to 43 region of F. Therefore, the F8(Mu)/FΔ(33-43) heteroduplexes consist of

duplex regions, due to homologous F sequences and the following nonhomologous regions. There is a substitution loop due to the bacterial DNA of F8 and the 8.5 to 16.3 region of F deleted in F8 (Sharp et al., 1972). There are two single-strand insertion-deletion loops, one due to the 33-43 kb sequence of F and one due to Mu prophage (Fig. 6). If Mu phage DNA containing a nonhomologous split end and a G-loop inversion is hybridized to this heteroduplex, the diheteroduplex structure depicted in Fig. 6a and 6b will be observed, depending on whether the split end sequence is or is not present in the prophage. In the three episomes studied, only the structure shown in Fig. 6b was observed (see Plate V). Therefore, in three independently derived Mu lysogens the split end sequence is absent in the prophage.

These results suggest that the split end sequences are random pieces of bacterial DNA picked up by Mu during its life cycle. They are lost during integration and are picked up again during induction of a lysogen.

ACKNOWLEDGMENT

We are grateful to Ellen Daniell for a gift of Mu phage.

REFERENCES

- Daniell, E., Abelson, J., Kim, J. S., and Davidson, N. (1973). Heteroduplex structures of bacteriophage Mu DNA. *Virology* 51, 237-239.
- Daniell, E., Boram, W., and Abelson, J. (1973). Genetic mapping of the inversion loop in bacteriophage Mu DNA. *Proc. Nat. Acad. Sci. U.S.* 70, 2153-2156.
- Hsu, M-T., and Davidson, N. (1972). Structure of inserted bacteriophage Mu-1 DNA and physical mapping of bacterial genes by Mu-1 DNA insertion. *Proc. Nat. Acad. Sci. U.S.* 69, 2823-2827.
- Sharp, P. A., Hsu, M-T., Ohtsubo, E., and Davidson, N. (1972). Electron microscope heteroduplex studies of sequence relations among plasmids of *E. coli*. I. Structure of F-prime factors. *J. Mol. Biol.* 71, 471-497.

LEGENDS TO THE FIGURES

Fig. 1a, b, and c. Structures seen in denatured and renatured Mu DNA samples, showing the two kinds of nonhomologous sequences, the G loop (Fig. 1b) and the split ends (Fig. 1a and 1b). Fig. 1c is the structure formed when the inverted G sequences of strands W_1 and W_2 base-pair (see Plate I). Fig. 1d depicts the sequences present in the two kinds of native molecule with either G or its inverted complement G' on the W strand.

Fig. 2. Structures seen in low formamide spreadings showing the presence of inverted repeat sequences on a single strand of Mu DNA. The top left sketch depicts the structure shown in Plate II. At the lower left, there is an enlarged drawing of the G loop part of this structure. The sequence organization responsible for this structure is depicted in the model at the right. The top drawing shows the sequences present in native Mu DNA at and near the G loop. One of the single strands produced by denaturation is shown. By base-pairing of the complementary sequences g_1 and g_1' and g_3 and g_3' , the model of the observed loop structure is formed.

Fig. 3. A model showing how reciprocal recombination of the inverted repeat sequences g_1 and g_1' results in inversion of the G loop region.

Fig. 4. (a). A structure of self-annealed Mu phage DNA showing the duplex regions formed between the g_3 sequences in the two arms of the G loop. It is observed in weakly denaturing condition. See also Plate IIIc. (b). A structure of self-annealed Mu prophage DNA under weakly denaturing condition. A micrograph of this structure is shown in Plate IIIId. (c) and (d). Structures of self-denatured Mu phage DNA with the inversion of the g_4 sequence.

Fig. 5. The left drawing depicts a self-renatured molecule from an episome bearing Mu prophage DNA, showing the G loop heterogeneity but no SE heterogeneity. The right drawing depicts a self-renatured molecule with both a G loop and heterogeneous split end sequences. The left molecule is seen in self-renatured preparations; the right one is not.

Fig. 6. Fig. 6c is a schematic representation of the F8(45F::Mu)/F Δ (33-43) heteroduplex. Two single-stranded insertion-deletion loops can be seen. The smaller one is the 33-43 kb sequence deleted in F Δ (33-43). The larger one is the DNA sequence of Mu prophage. Fig. 6a and 6b are the two possible structures of the F8(45::Mu)/F Δ (33-43)/Mu di-heteroduplex, with only the sequences around Mu prophage shown. Fig. 6a is the case when split end sequence is present in the prophage while in Fig. 6b the split end sequence is absent. Structure shown in Fig. 6b is the structure

actually observed. A corresponding micrograph is shown in Plate V.

LEGENDS TO THE PLATE

Plate I. Self-renatured Mu phage DNA showing the sequence inversion of the G loop. See Fig. 1c for a schematic model. The duplex region (G) has the same length as that of one arm of the G loop. A circular ϕ X174 single strand DNA, added as a length standard, is also shown.

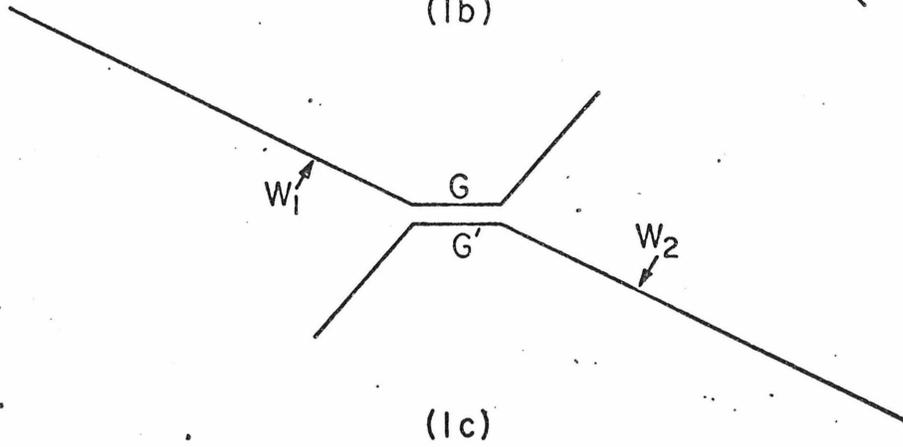
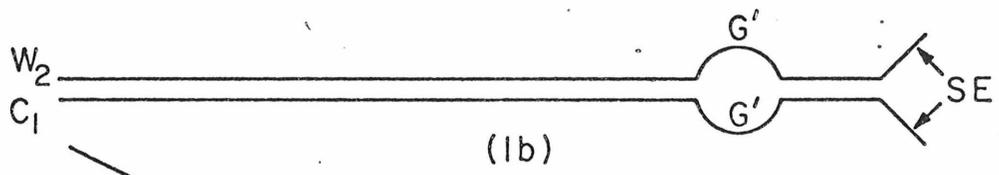
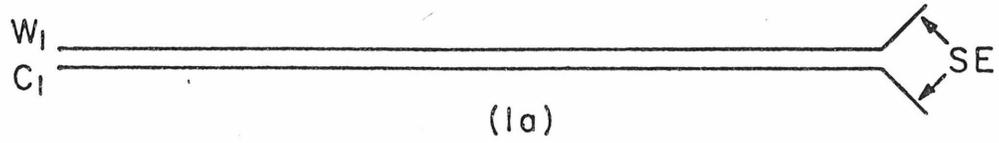
Plate II. Structure of a self-annealed, single-strand Mu phage DNA. The two inverted duplications are indicated by the arrows. The smaller inverted duplication is observed as a point. The two single-strand DNA segments between the two inverted duplications are constant in length and are equal to each other. For a schematic representation, see Fig. 2.

Plate III. (a) and (c), structures of the G loop region of self-renatured Mu phage DNA; (b) and (d), structures of the G loop region of self-renatured Mu prophage DNA carried in an episome, observed under weakly denaturing conditions. The longer inverted duplication seen in Plate II is present in the G loop of both Mu phage DNA (a) and Mu prophage DNA (b). The schematic representation of the structures shown in (c) and (d) are depicted in Fig. 4a and 4b, respectively.

Plate IV. Structure of the F8(45F::Mu)/F8-2 heteroduplex. The deletion loop of F8-2 is indicated by bs. The Mu prophage DNA in F8(45F::Mu) is seen as a single-strand

loop which is mapped against the F8-2 deletion.

Plate V. Structure of F8(45F::Mu)/F Δ (33-43)/Mu di-heteroduplex. A schematic representation is shown in Fig.5b.



<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>G</i>	<i>h</i>	(SE)	<i>W</i> ₁
<i>d'</i>	<i>b'</i>	<i>c'</i>	<i>d'</i>	<i>e'</i>	<i>f'</i>	<i>G'</i>	<i>h'</i>	(SE)'	<i>C</i> ₁

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>G'</i>	<i>h</i>	(SE)	<i>W</i> ₂
<i>d'</i>	<i>b'</i>	<i>c'</i>	<i>d'</i>	<i>e'</i>	<i>f'</i>	<i>G</i>	<i>h'</i>	(SE)'	<i>C</i> ₂

(1d)

Fig 1.

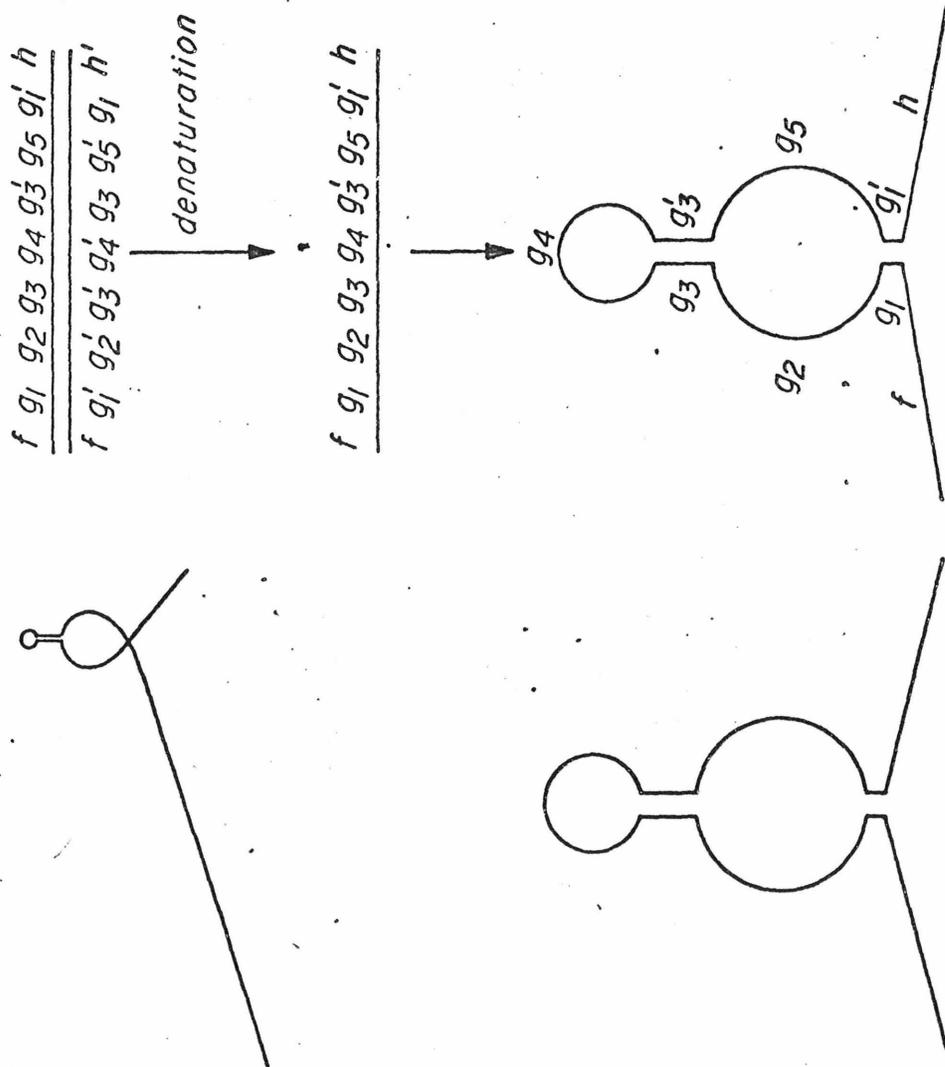


Fig. 2

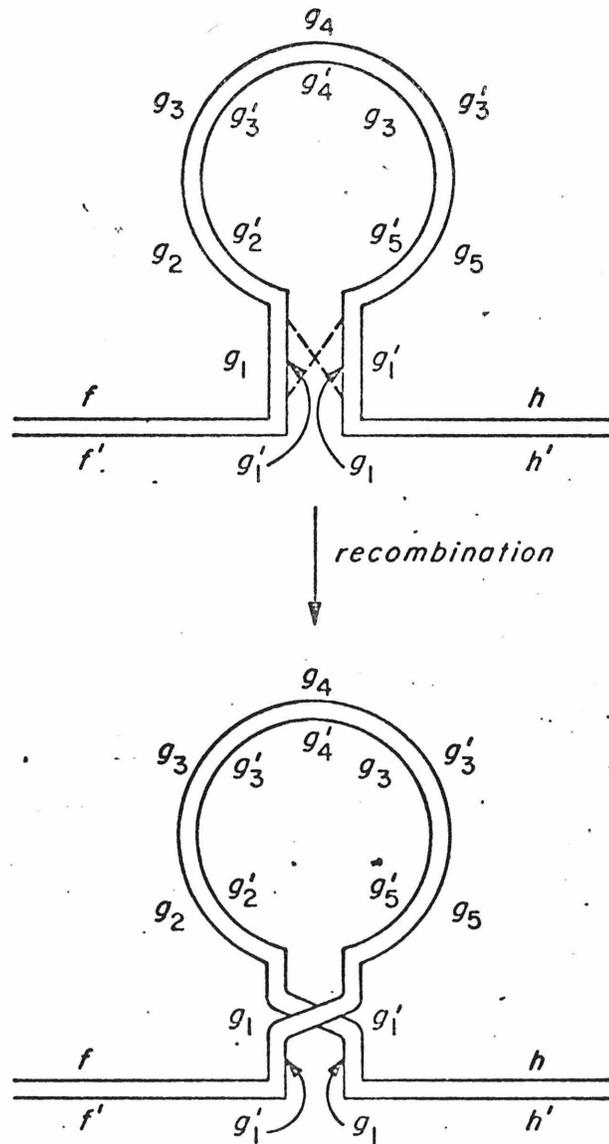


Fig. 3

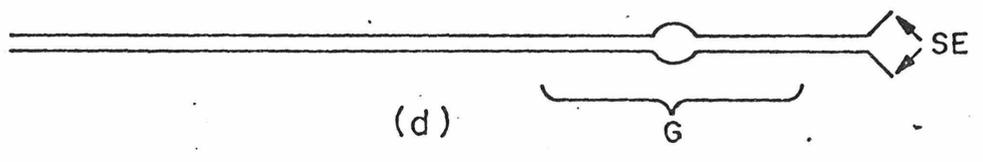
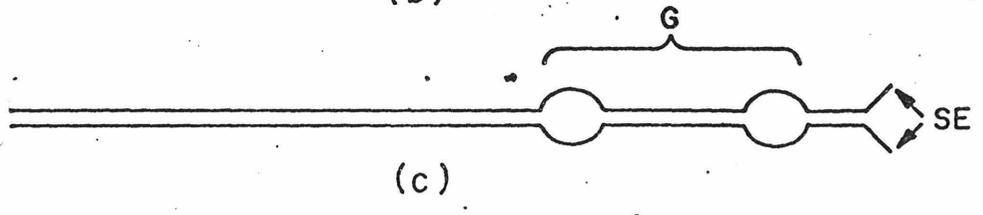
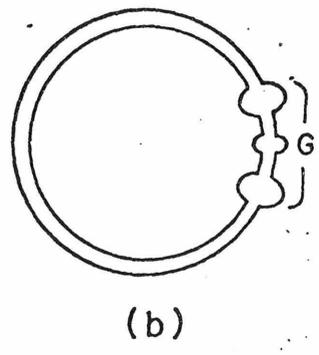
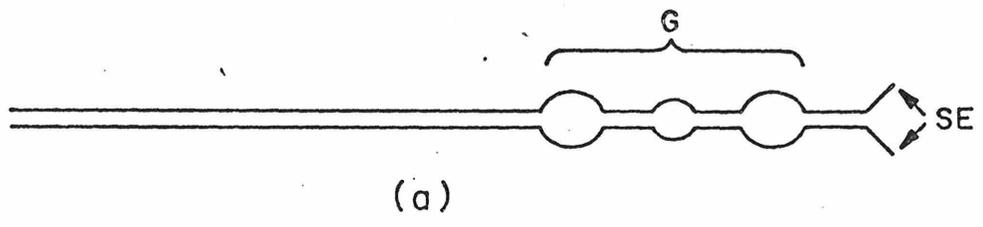


Fig 4

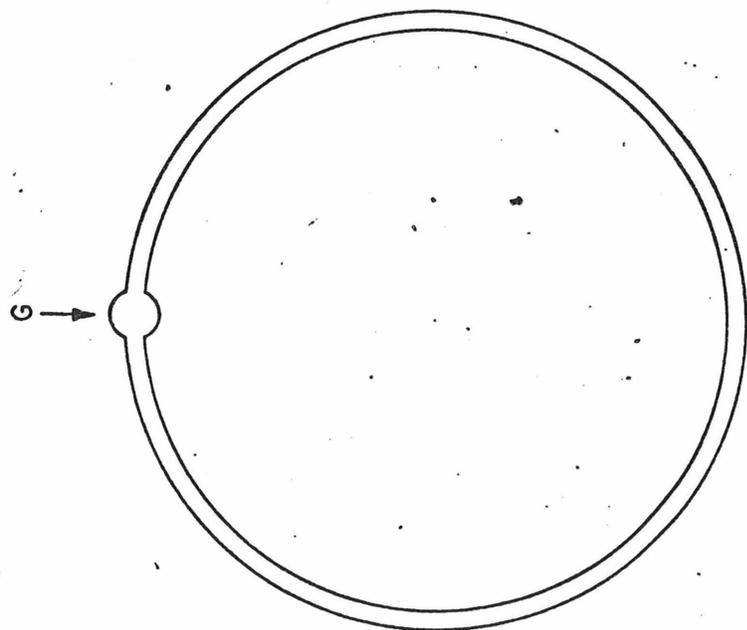
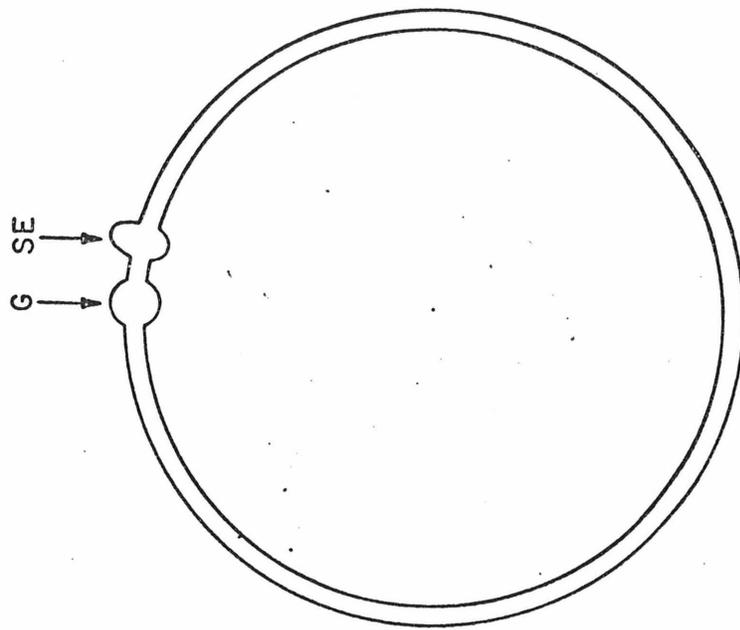
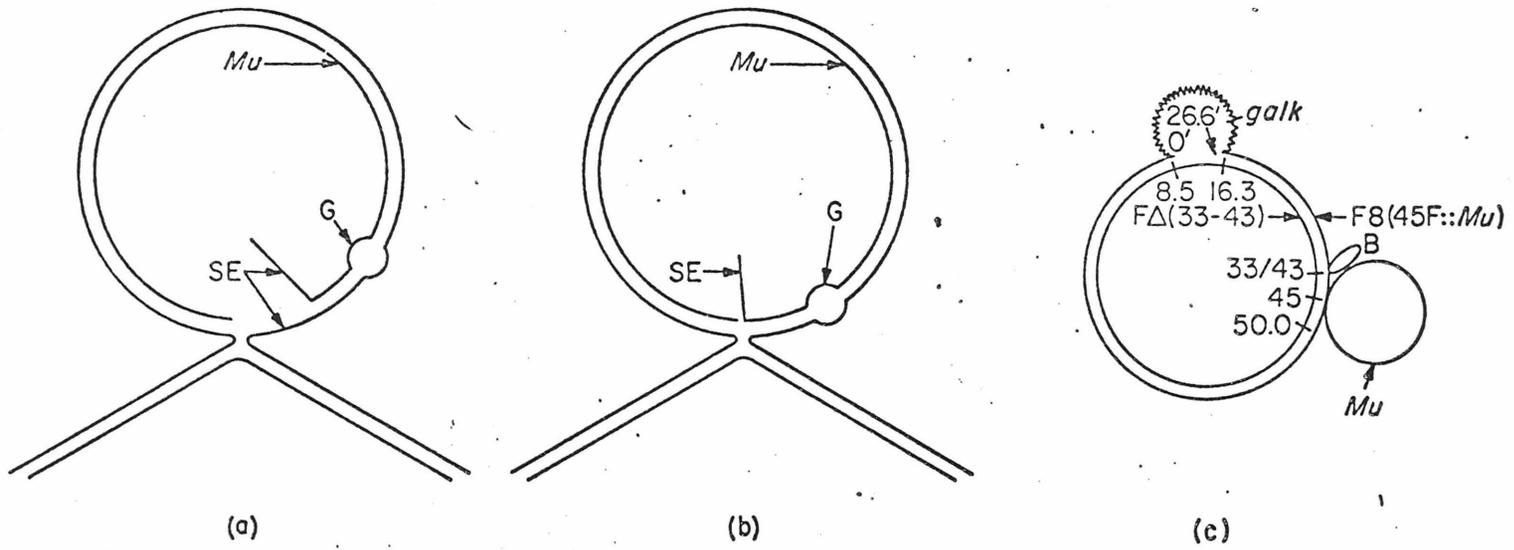


Fig. 5

Fig. 6



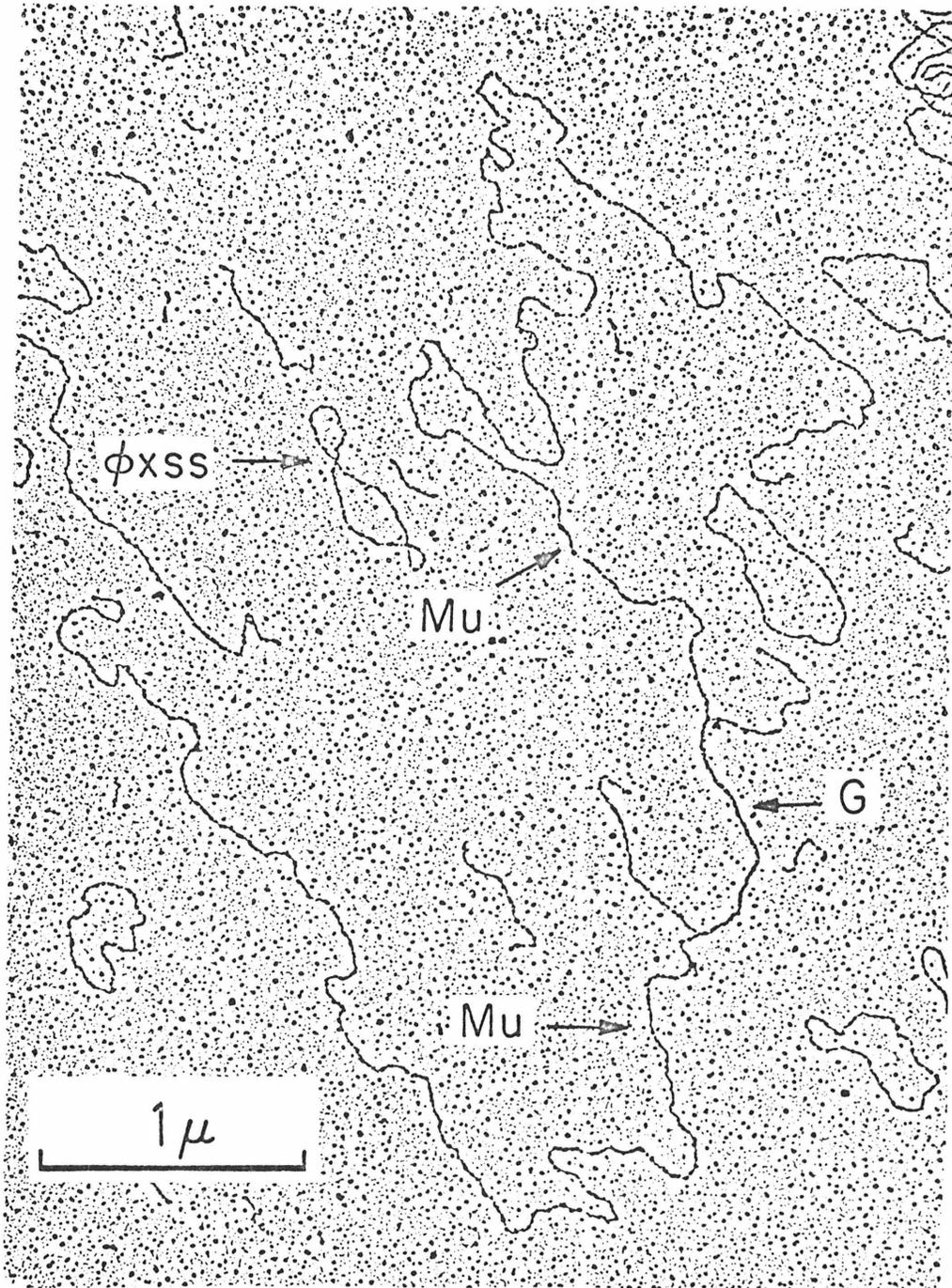


Plate I

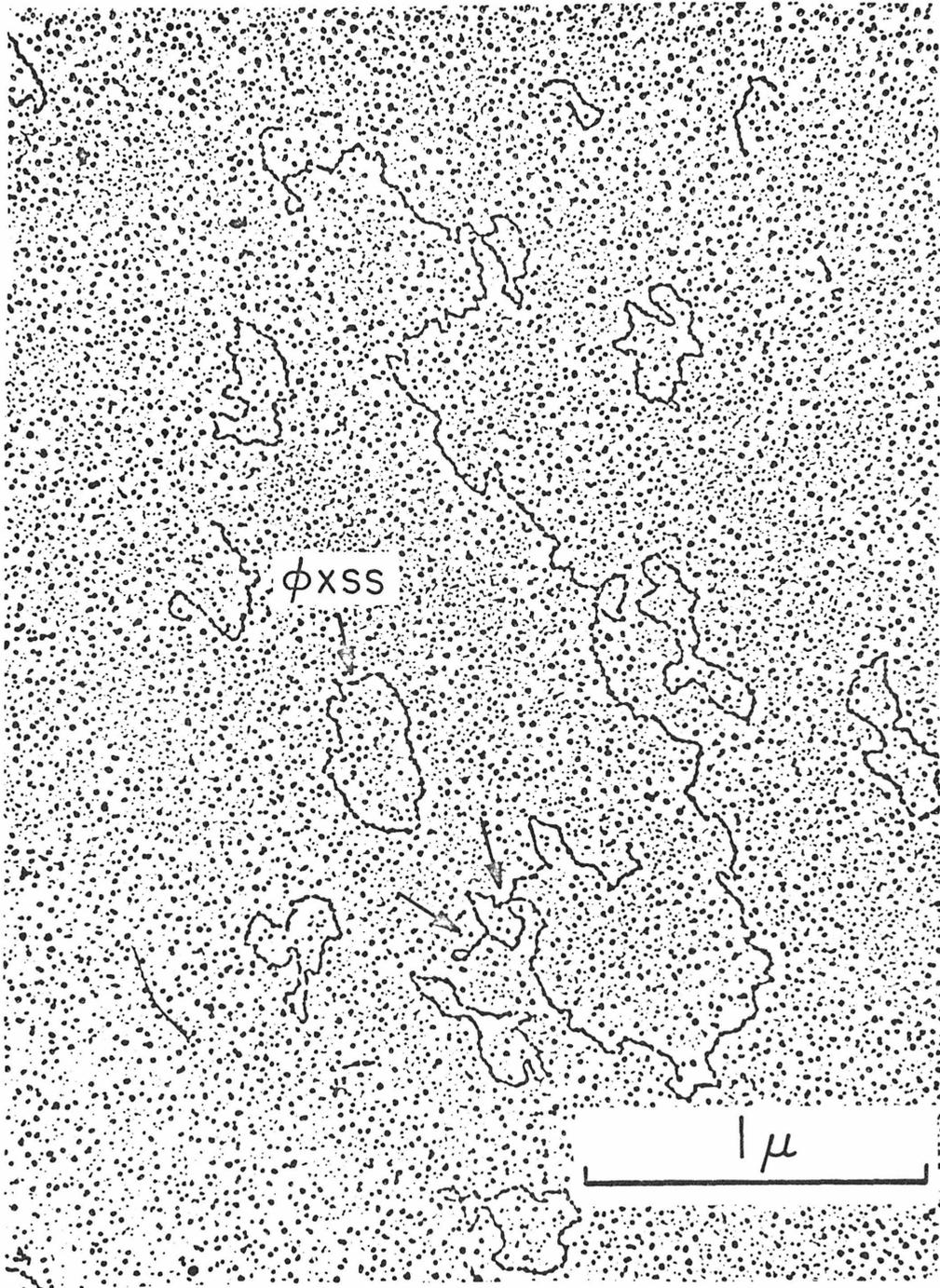
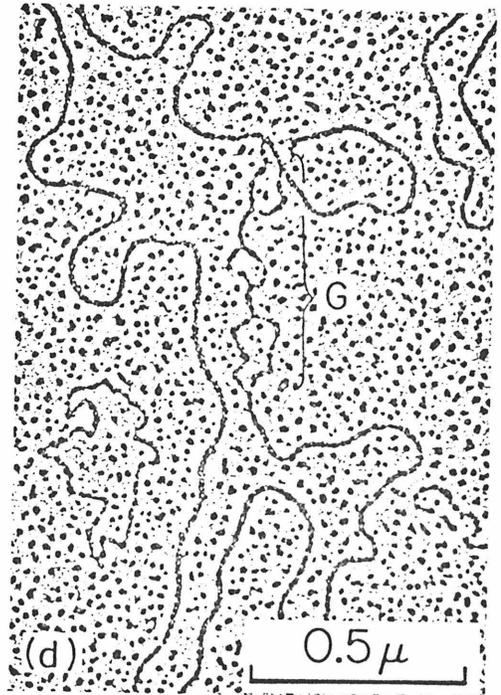
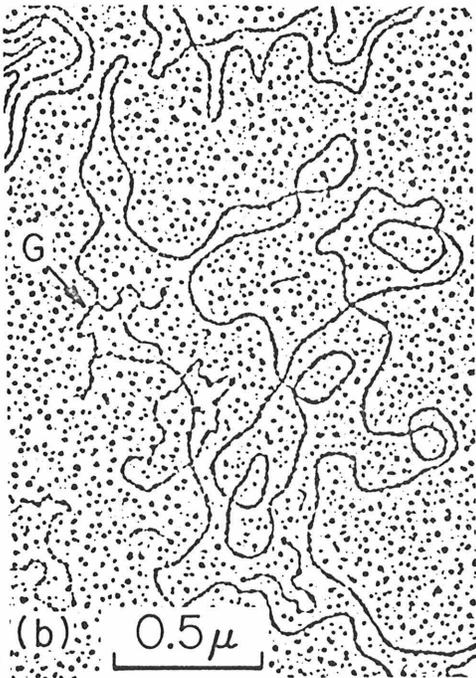
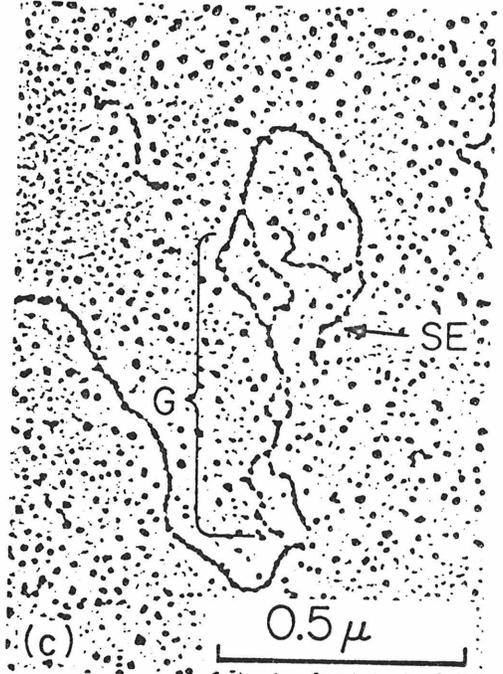
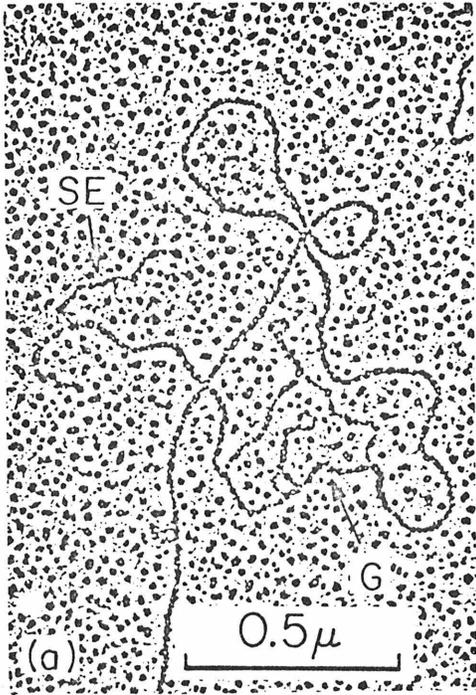


Plate II



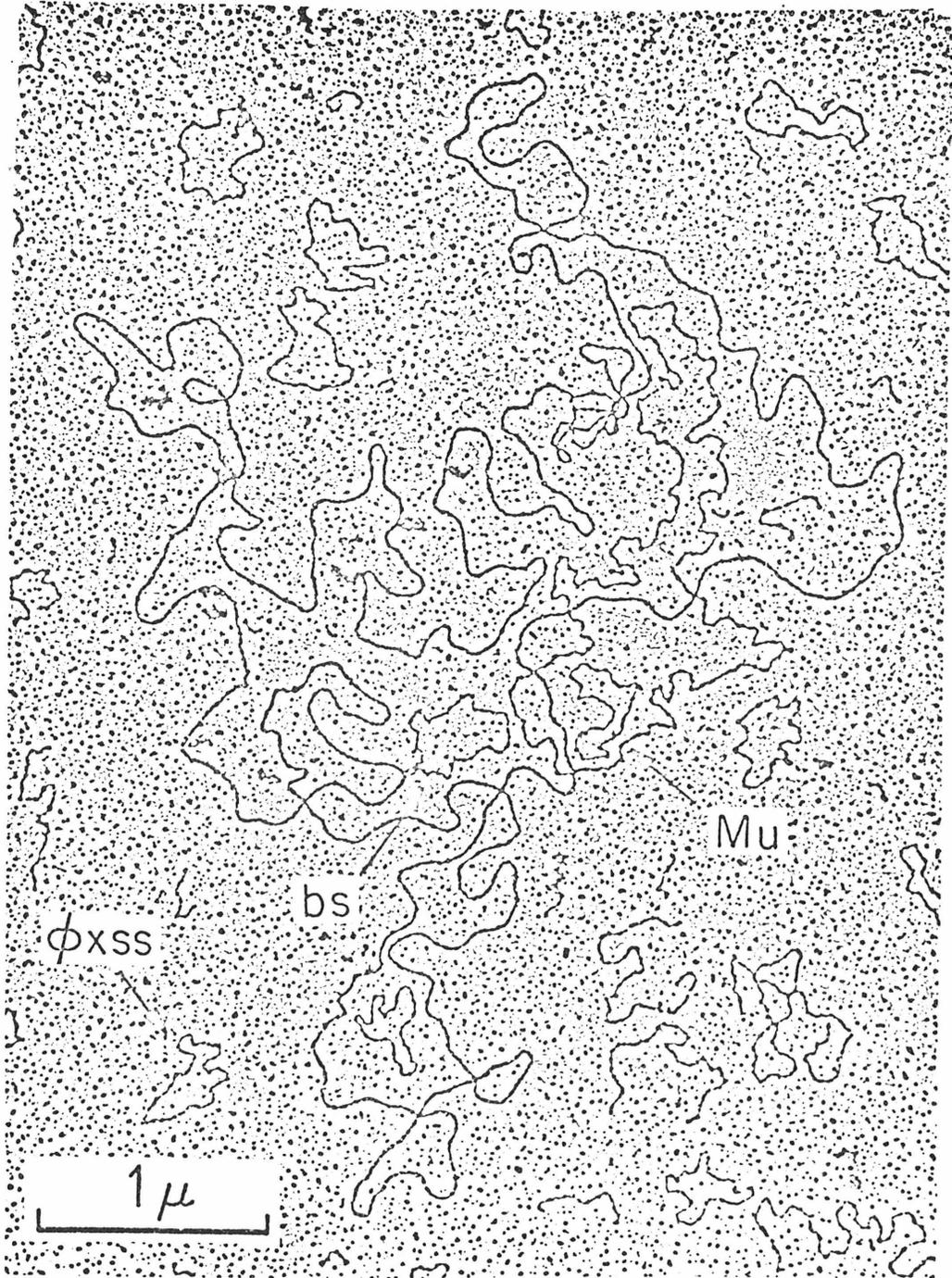


Plate IV

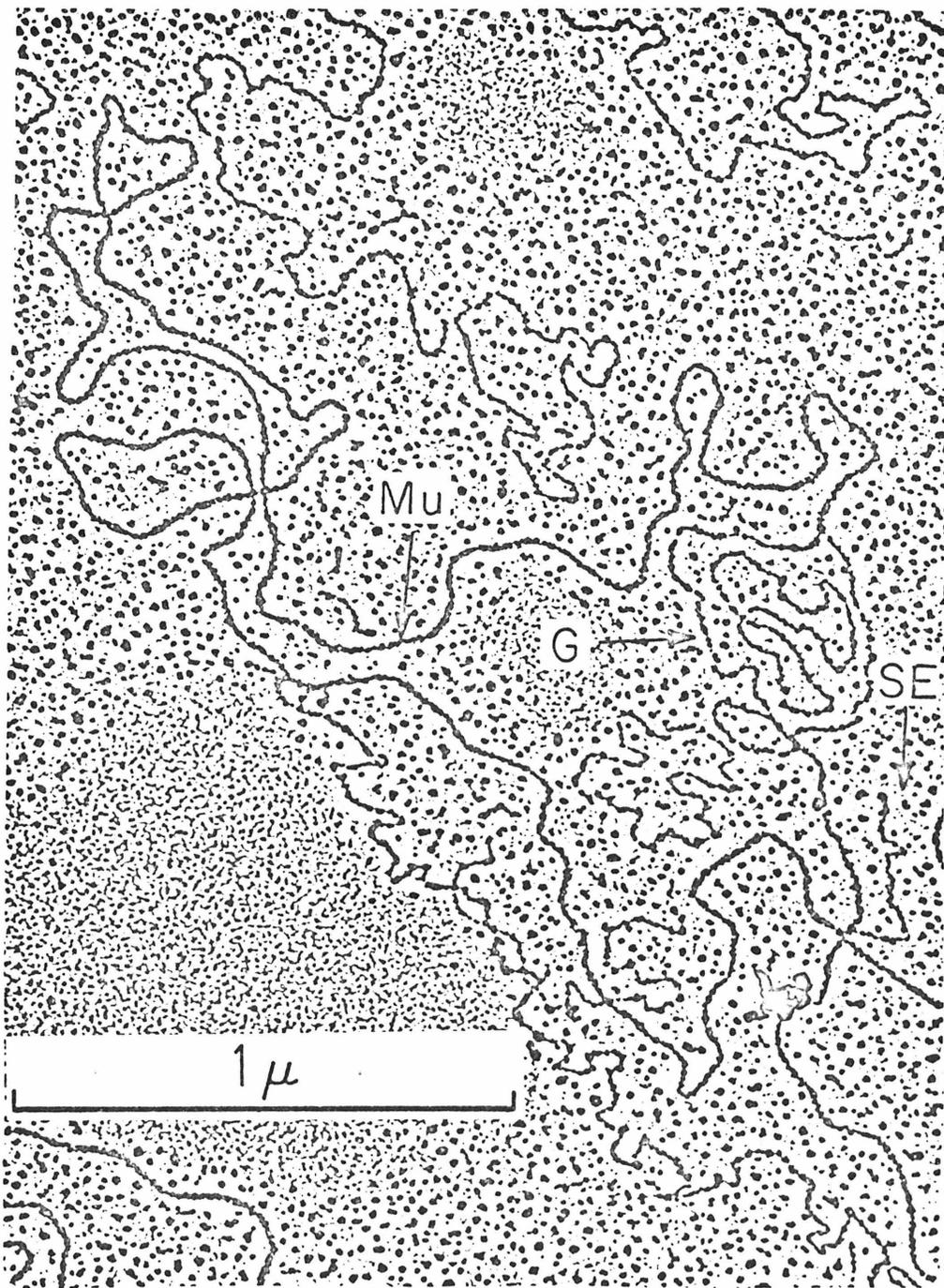


Plate V

Part III

AN ELECTRON MICROSCOPE STUDY OF
SINDBIS VIRUS RNA

This paper has been accepted for publication

by the

Cold Spring Harbor Symposia on

Quantitative Biology

An Electron Microscope Study of Sindbis Virus RNA

by

Ming-ta Hsu, Hsing-Jien Kung,

and Norman Davidson

///
Department of Chemistry
California Institute of Technology
Pasadena, California 91109

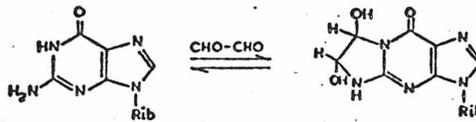
We wish to describe electron microscope methods for measuring the molecular lengths of RNA molecules and of mapping any poly-A sequences present. We chose to apply these methods initially to Sindbis virus RNA.

The virion of Sindbis virus contains only 42s RNA (Pfefferkorn *et al.*, 1967) with a molecular weight of $4.3 \pm 0.3 \times 10^6$ daltons, as estimated by DMSO sedimentation and gel electrophoresis (Simmons and Strauss, 1972). Sindbis RNA contains a poly-A sequence (Johnston and Bose, 1972 a, b). For most of the molecules its length is 60 to 80 nucleotides; for a small fraction (20% or less), the length is 150 to 250 nucleotides (Eaton and Faulkner, 1972).

In general, the basic protein film electron microscope techniques that work very well for both duplex and single-stranded DNA are less effective for RNA. The standard spreading conditions of 40 to 50% formamide and ca. 0.06 M ionic strength (Davis *et al.*, 1971) that cause single-strand DNA to be extended without denaturing duplex DNA do not fully extend most RNA species. RNA can be extended by spreading from a strongly denaturing solvent, in which most duplex nucleic acids would also be denatured (Granboulan and Scherrer, 1969; Robberson *et al.*, 1971; Verma *et al.*, 1970; Nanninga *et al.*, 1972). RNA molecules that have been cross linked by glutaraldehyde to gene 32 protein are well extended in standard formamide spreadings (Delius, Westphal, and Axelrod, 1973). This treatment is moderately denaturing for duplex nucleic acids which are rich in A-T or A-U base pairs.

We decided to investigate glyoxal as a reagent that would block some of the hydrogen bonding functions of a nucleic acid and thus increase the ease with which the molecule could be extended, without

affecting the stability of A-T and A-U base pairs. Glyoxal adds to adenosine and cytidine rapidly and reversibly with the rather small equilibrium association constants of 1.4 M^{-1} and 5.8 M^{-1} respectively at 20°C . It reacts more slowly with guanosine. The equilibrium association constant is about $6 \times 10^3 \text{ M}^{-1}$. The half-life of the glyoxal-guanosine adduct is directly proportional to the hydrogen ion concentration and is about 50 hours at 20° and pH 7 (Broude and Budowsky, 1971). The probable structure of the adduct is indicated below.



Under proper conditions therefore it should be possible to block the hydrogen bonding functionalities of the G residues of a nucleic acid without interfering with the ability of a poly-A sequence to hybridize to an added poly-dT strand. It seemed to us that this might be the basis of an electron microscope method of mapping the positions of poly-A sequences on RNA molecules.

RESULTS

a) Sedimentation coefficients and length measurements of glyoxal treated RNA. E. coli 16s and 23s rRNA and Sindbis virus RNA were prepared as described in the legend to Fig. 1. The sedimentation profiles in Fig. 1 show that in neutral sucrose, the Sindbis virus RNA preparation has a sedimentation coefficient of 43s, by

comparison with E. coli 23s rRNA.

The RNA components isolated from the gradients were treated for 1 hr with 0.5 M glyoxal at 37°C in 0.01 M phosphate buffer at pH 7. It was thought that the low electrolyte concentration would minimize the amount of secondary structure in the RNA's thus increasing the reactivity with glyoxal.

The reaction products were sedimented through a sucrose gradient containing 0.1 M glyoxal. These results are also shown in Fig. 1. The sedimentation velocities of the several RNA species are all reduced by a factor of about two on treatment with glyoxal, thus indicating that some of the secondary structure is disrupted.

With the glyoxal concentration present in the sucrose gradient, a moderate fraction of the A and C residues and all of the G residues should be present as glyoxal adducts, according to the equilibrium data quoted above.

When RNA treated with glyoxal as described above is spread from 50% or 40% formamide in the basic protein film procedure, the molecules appear as extended thin threads. Typical micrographs are shown in Fig. 2. The samples are well extended in spreadings from lower formamide concentrations or even in aqueous spreadings (no formamide); in the latter case, the molecular lengths are about 60% of those from 40 or 50% formamide. Histograms of the length distributions are displayed in Fig. 3. It may be seen that the procedure gives a reasonably narrow length distribution curve for a homogeneous RNA. The samples are diluted by a factor of about 10^2 from

the glyoxal reaction mixture into the spreading solution. There is tris buffer in the spreading solution. Either of these factors alone should cause the removal of glyoxal from A and C residues. Therefore, in the molecules seen in the electron micrographs, it is probable that most of the G residue, and only the G residues, are present as glyoxal adducts.

Table 1 summarizes the length of E. coli rRNA molecules as spread by several procedures. There are small differences between the lengths as measured by the several procedures. Our value for 23s rRNA, in conjunction with a molecular weight of 1.08×10^6 (Attardi and Amaldi, 1970), and a residue weight for E. coli Na rRNA of 345, gives an internucleotide spacing of 2.5_6 \AA . Delius et al. (1972) find an internucleotide spacing of 3.7 \AA for T7 mRNA by the gene 32 procedure.

The molecular lengths of Sindbis and 23s rRNA are measured as $3.4_5 \pm 0.14 \mu$ and $0.80 \pm 0.05 \mu$, respectively. If the molecular weight of the latter is taken as 1.08×10^6 daltons (sodium salt), the molecular weight of Sindbis RNA is calculated as $4.6_5 \pm 0.4 \times 10^6$. Simmons and Strauss (1972) give $4.3 \pm 0.3 \times 10^6$.

Finally, we wish to mention that treatment with 0.5 M glyoxal as described above does not denature duplex DNA nor does it cause any opening up of the b2 deletion loop in a λ/λ_{b2b5} heteroduplex (A. Forsheit, personal communication). Therefore, glyoxal may be used to extend single-strand DNA or RNA without affecting long, perfect duplex regions.

b) Sindbis virus RNA has cohesive ends. When Sindbis virus RNA is treated with glyoxal under the extreme conditions described above, almost all of the molecules seen are linear. We have observed that when Sindbis RNA is subjected to any one of several more gentle denaturing treatments, which are sufficient to cause partial extension of the molecules, about 25% to 50% (depending on the preparation) of the RNA molecules in an electron microscope grid are circular, with a duplex "handle" at one point on the circle.

For example, if Sindbis RNA in 0.010 M phosphate buffer, pH 8.0, is treated with 0.010 M glyoxal at 37°C, and then (after ca. 10^2 -fold dilution) spread from 30% formamide, 0.10 M tris, 0.010 M EDTA, pH 8.5 onto a hypophase of 5% formamide with one-tenth the electrolyte concentration, the following observations are made as a function of time of incubation with glyoxal. At zero time, the molecules are collapsed bushes. After 20 minutes, they are slightly extended and many are circular. After 75 minutes, the molecules are more extended, and 25 to 50% of the molecules are scored as circles with an apparent handle or duplex stem. The stem has a length of about 0.083μ corresponding to 250 ± 50 nucleotide pairs. Typical micrographs are shown in Fig. 4. The molecules consist of smooth regions, secondary structure bumps of variable position and appearance, and the handle. The measured single-strand contour length, treating all secondary structure features as duplex is $2.76 \pm 0.10\mu$. Thus the molecules appear to be less extended than those shown in

Fig. 2. On still longer treatment with glyoxal, the circular structures open up, before all of the secondary structure is disrupted, and the molecules become as extended and smooth as those shown in Fig. 2.

If Sindbis RNA is spread from 40% formamide, 0.10 M tris, 0.010 M EDTA, pH 8.5, without treatment with glyoxal, 10% to 50% of the molecules are circular, with considerable secondary structure, and with a handle. Electron micrographs are displayed in Fig. 4. If 50% formamide is used, most of the circles have opened up, and the RNA is smoother and more extended. No circles are seen in spreadings from 60% formamide. It is our impression that the 40% formamide solvent represents a delicate balance between the denaturing power needed to partially extend the RNA and that which will cause opening of the circles. The variation of frequency of circles in different experiments may be largely due to small variations in spreading conditions.

Under proper annealing conditions, linear molecules with undamaged ends can reform circles. Thus, if Sindbis RNA is dialyzed against pure formamide at 4°C for one to two hours, and diluted into a 40% formamide spreading solution, the molecules are all linear. If, after dissociation by pure formamide, the RNA molecules are dialyzed into 40% formamide, 0.20 M phosphate buffer, pH 8.0 ($[Na^+] = 0.38 \text{ M}$) and incubated at 4°C, and then diluted into the 40% formamide spreading solution, the fraction of the long (and therefore possibly intact) molecules which are circular rises over a period of 4

hours to about 25%. Longer incubations are not useful because of degradation of the RNA.

c) Sindbis RNA has a poly-A sequence at one end. We chose to use poly-dT as a complementary polynucleotide strand to physically label the poly-A sequence of Sindbis RNA for EM mapping. It has several advantages for this purpose. Poly-dT is available as a high molecular weight polymer from General Biochemicals. The rA:dT duplex melts about 7° higher than the rA:rU duplex (Riley *et al.*, 1966). Poly-dT itself gives an extended filament, even from our aqueous spreading solution.

The available evidence suggests that poly-A sequences, if present, are at the 3' end of RNA molecules. A terminal poly-A sequence could base pair with a stretch of T residues anywhere along the poly-dT chain. The complex would then have a Y-shaped structure, with one branch being Sindbis RNA and two branches poly-dT. If there were a poly-A sequence in the middle of an RNA molecule, the complex would have an X shape. A unique interpretation of a Y-shaped structure, for example, is complicated by the facts that many of the Sindbis RNA molecules are fragmented and that the poly-dT has a heterogeneous length distribution. However, it is possible to distinguish between Sindbis RNA and poly-dT by spreading under conditions in which the former has some residual structure whereas the latter does not.

By studying complexes of pure poly-rA with poly-dT, we observed that in our standard electrolyte, the strands are mainly

associated as duplexes in 40% formamide but significantly dissociated into single strands in 50% formamide. This observation is reasonably consistent with the report by Riley *et al.* (1966) that the rA:dT complex melts at 64° in 0.15 M Na⁺.

Hybridization of poly-dT with Sindbis RNA was carried out by denaturing the mixture with pure formamide, renaturing at 37°C in 0.1 M NaCl, 0.01 M tris, 0.001 M EDTA, pH 7.0 for several hours. Each polynucleotide concentration was about 100 µg/ml. The samples were diluted and spread from 40% formamide. A large number of Y-shaped structures were observed. The RNA branch can be distinguished from the two dT branches because the former has secondary structure knobs and, even where it appears extended, is thicker. Typical micrographs are shown in Fig. 5.

Alternatively, Sindbis RNA was treated with 0.5 M glyoxal as described above. It was dialyzed against 0.10 M phosphate buffer, pH 8.0, at 37°C for three hours. Poly-dT was added and the mixture incubated at 37° for several hours. It was then spread from 30% or 20% formamide (hypophase, 5% and 0% formamide). Typical micrographs are shown in Fig. 6. In 30% formamide, the distinction between glyoxal treated Sindbis RNA and dT is rather subtle; the RNA is thicker and smoother. In 20% formamide, the RNA is more curved and shows some secondary structure knobs. The structures seen are Y-shaped. We conclude that the poly-A sequence on Sindbis RNA is at one end.

A small number of circular Sindbis RNA molecules with dT attached to the handle were seen. An example is shown in Fig. 5.

This result suggests that the poly-A sequence of Sindbis RNA is not at all or only partially involved in the base pairing of the cohesive ends. In this connection, we wish to mention that when Sindbis RNA is incubated at 4° in 0.20 M phosphate buffer, 40% formamide, as described for the reannealing of the cohesive ends, but in the presence of poly-dT at a concentration of 150 µg/ml, the frequency of circle formation was reduced from 26% to 7%. The interpretation of this observation is not certain however, because controls as to possible nuclease effects or nonspecific interactions were not done.

DISCUSSION

Several techniques are now available for extending RNA for electron microscope studies. Different methods will probably be useful for different purposes. Glyoxal procedures are simple and versatile and deserve further study. It may be mentioned that in our experience E. coli rRNA is just as difficult to spread in an extended form as is Sindbis RNA. However, Sindbis RNA has a higher T_m and a more cooperative transition than many other RNA's, including E. coli rRNA, TMV, R17, Rauscher virus RNA (Sprecher-Goldberg, 1966). The correlation between thermal denaturation parameters and EM spreading characteristics for RNA's is evidently not simple.

The most unexpected result reported here is the presence of circular molecules in the preparation of Sindbis virus RNA extracted from the virion and the fact that the cohering ends are capable of reversible dissociation and reassociation. The data do not exclude the possibility that all the native molecules within the virions are circular,

and the linear molecules observed are due to degradative effects. The RNA had been deproteinized with phenol and SDS. It is unlikely that the cohesion of the ends is due to an attached protein.

The observed handle in the circular molecule could be a secondary structure feature in the middle of the linear molecule. We believe that this is incorrect and that the handle is due to the cohesive sequences very close to the ends. The observation of circular molecules with poly-dT attached to the handle (Fig. 5) supports this interpretation. Furthermore, we see a very low frequency of circles in Sindbis RNA after the 0.5 M glyoxal treatment. These molecules always have a handle. The linear molecules are smooth with no handle or other secondary structure features. Thus the handle is due to the duplex formation leading to cyclization. The handle structure is consistent with inverted repeat sequences close to the ends that are paired in a standard antiparallel Watson-Crick structure. We have estimated that the length of the handle corresponds to 250 ± 50 base pairs. Evidently, one of the cohesive sequences is not quite at the end of the molecule, because there is a poly-A sequence beyond it.

It may be recalled that the single strands of several types of adeno virus and of adeno-associated virus have cohesive ends (Garon *et al.*, 1972; Wolfson and Dressler, 1972; Koczot *et al.*, 1973).

In the case of Sindbis RNA, the circular structure with the handle may be significant for replication and/or translation.

ACKNOWLEDGMENT

This research has been supported by NIH grant GM-1991. We are deeply grateful to Professor James Strauss and Dr. Ellen Strauss for generous assistance and loan of facilities for growing Sindbis virus, and their counsel as to the preparation and properties of the RNA.

TABLE 1

rRNA Length Measurements^a by Several Spreading Procedures

Method	16s <u>E. coli</u>	23s <u>E. coli</u>	18s HeLa
Urea ^b	0.38 ± 0.03	0.79 ± 0.08	0.59 ± 0.04
Urea ^c	0.40 ± 0.01	0.72 ± 0.02	
Formamide/urea ^d			0.51 ± 0.06
Dimethyl sulphoxide ^e	0.43	0.84	
Formamide/glyoxal ^f	0.43 ± 0.04	0.80 ± 0.05	

a) Number average lengths and standard deviations, all in μm .

b) Granboulan and Scherffer (1969).

c) Verma et al. (1970).

d) Robberson et al. (1971).

e) Nanninga et al. (1972).

f) This paper.

REFERENCES

- ATTARDI, G., and F. AMALDI. 1970. Structure and synthesis of ribosomal RNA. *Ann Rev. Biochem.* 39:183.
- BROUDE, N. E., and E. I. BUDOWSKY. 1971. The reaction of glyoxal with nucleic acid components. III. Kinetics of the reaction with monomers. *Biochim. Biophys. Acta* 254: 380.
- DAVIS, R. W., and R. W. HYMAN. 1971. A study in evolution: the base sequence homology between coliphages T7 and T3. *J. Mol. Biol.* 61: 287.
- DAVIS, R. W., M. SIMON, and N. DAVIDSON. 1971. Electron microscopic heteroduplex methods for mapping regions of base sequence homology in nucleic acids. In (L. Grossman and K. Moldave, eds.) *Methods in Enzymol.* Academic Press, New York XXI, p. 413.
- DELIUS, H., H. WESTPHAL, and N. AXELROD. 1973. Length measurements of RNA synthesized in vitro by E. coli RNA Polymerase. *J. Mol. Biol.* 74: 677.
- EATON, B. T., and P. FAULKNER. 1972. Heterogeneity in the poly-A content of the genome of Sindbis virus. *Virology* 50: 865.
- GARON, C. F., K. W. BERRY, and J. A. ROSE. 1972. A unique form of terminal redundancy in adenovirus DNA molecules. *Proc. Nat. Acad. Sci. USA* 69: 2391.
- Granboulan, N., and K. SCHERRER. 1919. Visualization in the electron microscope and size of RNA from animal cells. *Europ. J. Biochem.* 9: 1.

REFERENCES (continued)

- JOHNSTON, R. E., and H. R. BOSE. 1972a. Adenylate-rich segment in the virion RNA of Sindbis virus. *Biochem. Biophys. Res. Commun.* 46: 712.
- JOHNSTON, R. E., and H. R. BOSE. 1972b. Correlation of messenger function with adenylate-rich segments in the genomes of single-stranded RNA viruses. *Proc. Nat. Acad. Sci. USA* 69: 1514.
- KOCZOT, F. J., H. J. CARTER, C. F. GARON, and J. A. ROSE. 1973. Self-complementarity of terminal sequences within plus or minus strands of adenovirus-associated virus DNA. *Proc. Nat. Acad. Sci. USA* 70: 215.
- NANNINGA, N., M. MEYER, P. STOOFF, and REIJNDERS, I. 1972. Electron microscopy of *Escherichia coli* ribosomal RNA: spreading without a basic protein film. *J. Mol. Biol.* 72: 807.
- PFEFFERKORN, E. R., B. W. BURGE, and H. M. COADY, 1967. Intracellular conversion of the RNA of Sindbis virus to a double-stranded form. *Virology* 33: 239.
- RILEY, M., B. MALING, and M. J. CHAMBERLIN. 1966. Physical and chemical characterization of two- and three-stranded adenine-thymine and adenine-uracil homopolymer complexes. *J. Mol. Biol.* 20: 359.

REFERENCES (continued)

- ROBBERSON, D., Y. ALONI, G. ATTARDI, and N. DAVIDSON. 1971. Expression of the mitochondrial genome in HeLa cells. XI. Size determination of mitochondrial ribosomal RNA by electron microscopy. *J. Mol. Biol.* 60: 473
- SIMMONS, D. T., and J. H. STRAUSS. 1972. Replication of Sindbis virus. I. Relative size and genetic content of 26s and 49s RNA. *J. Mol. Biol.* 71: 599.
- SPRECHER-GOLDBERGER, S. 1966. Differences between the structures of poliovirus and Sindbis virus infectious ribonucleic acids. *Arch. Gesamte Virusforsch.* 20: 225.
- VERMA, I. M., M. EDELMAN, M. HERZBERG, and U. Z. LITTAUER. 1970. Size determination of mitochondrial ribosomal RNA from Aspergillus nidulans by electron microscopy. *J. Mol. Biol.* 52: 137.
- WOLFSON, J., and D. DRESSLER. 1972. Adenovirus-2 DNA contains an inverted terminal repetition. *Proc. Nat. Acad. Sci. USA* 69: 3054.

LEGENDS TO FIGURES

Fig. 1. Sedimentation profiles of Sindbis RNA and of E. coli rRNA. (a) Sedimentation through a 5 to 30% sucrose gradient in 0.01 M phosphate buffer, pH 7, 0.05% SDS at 15°C, 44 K rpm, 0.5 hrs, SW 50.1 rotor; 0.1 ml of 100 µg/ml ³H Sindbis RNA was layered on top of the gradient. Slightly more ³H E. coli rRNA was run in a separate tube at the same time. (b) A mixture of ³H Sindbis RNA and ¹⁴C E. coli rRNA was treated with glyoxal as described in the text and sedimented through the same gradient as in (a) but containing 0.1 M glyoxal, same sedimentation conditions, but 5.0 hrs. The velocity ratios between (a) and (b) are: Sindbis RNA, 2.3; 23s, 1.8; 16s, 1.4.

Sindbis virus was grown in a primary culture of chick embryo fibroblasts prepared by Dr. J. Strauss. His procedure (personal communication) for growth and harvest of the virus was used. Virions were purified by sucrose gradient velocity centrifugation and then pelleted. RNA was purified by phenol extraction at 4°C using ^{double}distilled water saturated phenol containing 0.1% of hydroxy quinoline. The aqueous phase was 0.1 M NaCl, 0.01 M tris, 0.001 M EDTA, pH 7.0, 1.0% SDS. Degradation of the RNA is less if phenol is added to the virus pellet before adding the aqueous SDS buffer (P. Vogt, J. Strauss, private communications).

¹⁴C or ³H labeled E. coli rRNA was extracted from ribosomes (a gift from Drs. R. Deonier and A. Forsheit) with phenol-SDS as described above.

Fig. 2. Electron micrographs of Sindbis RNA and E. coli 23s rRNA after treatment with glyoxal as described in the text. After reaction or isolation from the 0.1 M glyoxal sucrose gradients, the samples were diluted ca. 10^2 fold into 50% formamide, 0.1 M tris, 0.01 M EDTA, pH 8.5 and spread onto 20% formamide with one-tenth the electrolyte concentration (Davis et al., 1971). The relation between the formamide concentration in the hyperphase and the hypophase was: 60% to 30%, 50% to 20%, 40% to 10%, and 30% to 5% (Davis and Hyman, 1971).

Fig. 3. Histograms of length distributions of Sindbis RNA and E. coli rRNA after treatment with glyoxal and EM spreading from 50% formamide as described in the legend to Fig. 2. Essentially the same distribution was obtained by direct dilution of the 0.5 M glyoxal reaction mixture or with the peak samples from the 0.1 M glyoxal sucrose gradient centrifugation, but there is more short degraded material in the former case.

Fig. 4. Electron micrographs of Sindbis RNA, treated as described in the text, showing circular molecules. In several cases, the duplex handle is identified with an arrow. The top two micrographs are after treatment with 0.010 M glyoxal for 75 minutes, and spread from 30% formamide; the bottom one is a spreading from 40% formamide without prior glyoxal treatment. For other details about the spreading solution and hypophase, see legend to Fig. 2.

Fig. 5. Poly-dT hybridized to the ends of Sindbis RNA in 40% formamide spreadings without glyoxal as described in the text. The

bottom micrograph shows a circular RNA molecule attached at the handle to poly-dT. The poly-dT had a length distribution ranging from 1 to 6 μm . It was obtained from the General Biochemicals Company, Chagrin Falls, Ohio 44022.

Fig. 6. Poly-dT hybridized to the ends of Sindbis RNA after the RNA has been treated with glyoxal as described in the text. The two top micrographs are for spreadings from 30% formamide. The bottom one is a spreading from 20% formamide. The top left micrograph shows one dT molecule attached to the ends of two Sindbis RNA molecules.

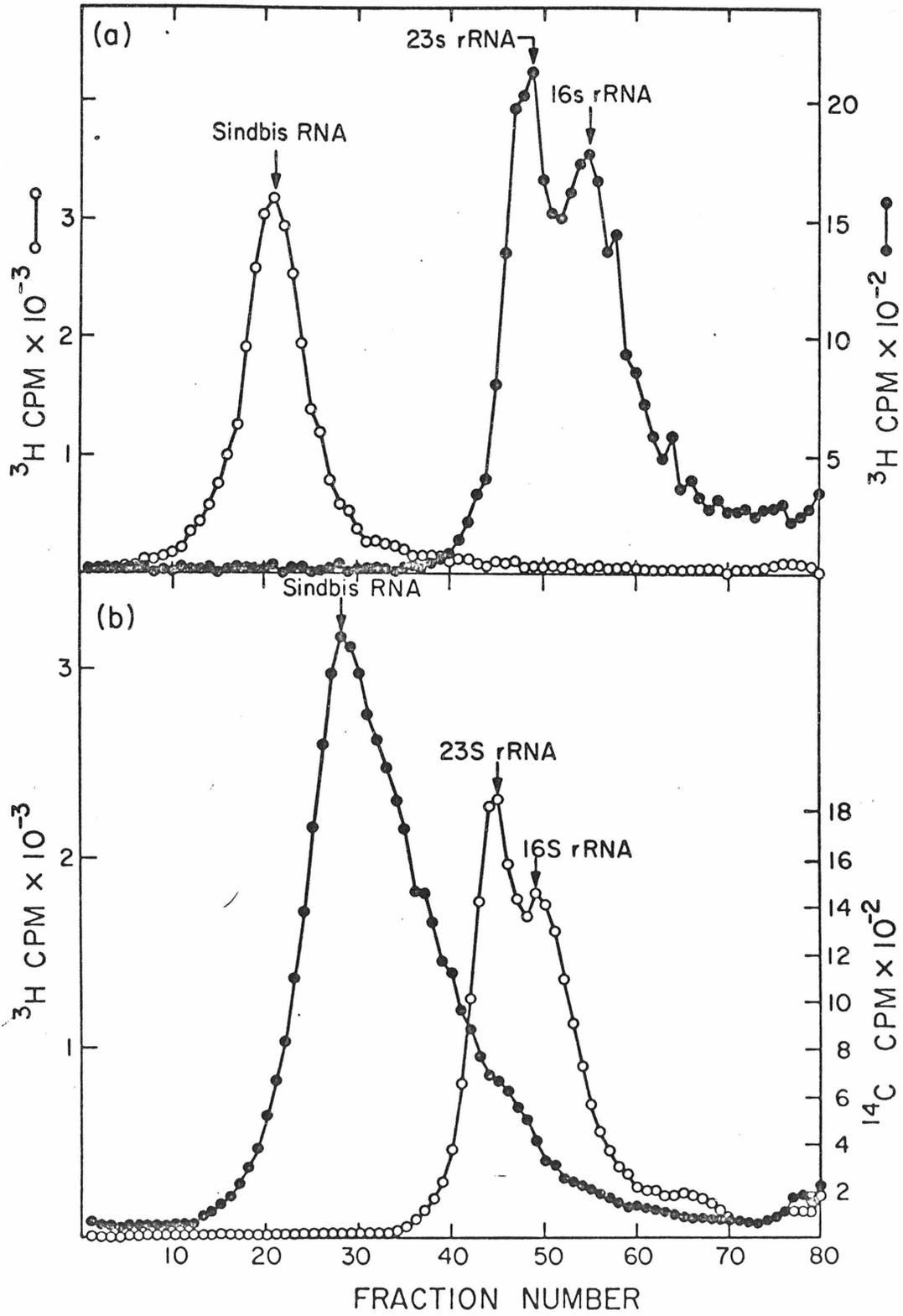


Figure 1

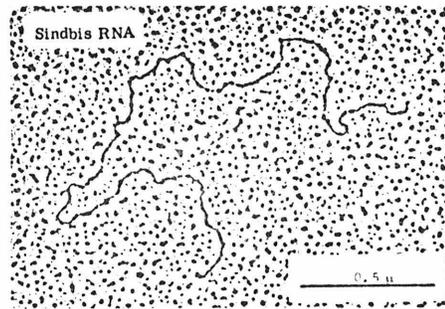


Figure 2

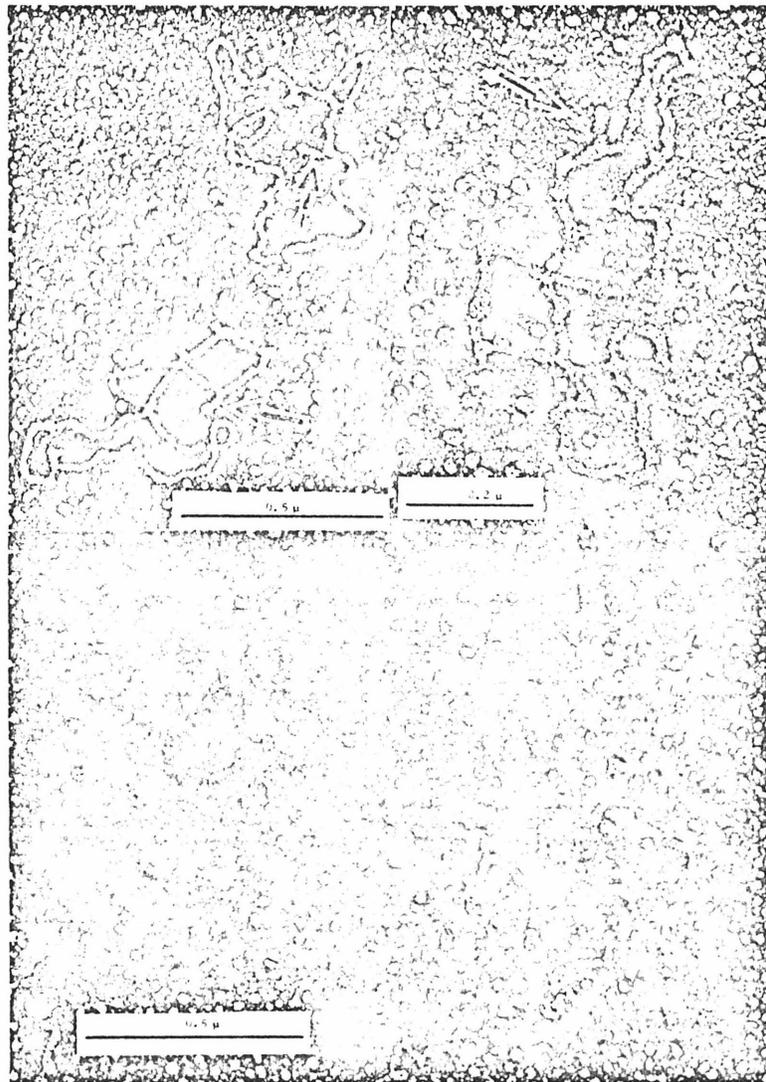


Figure 4

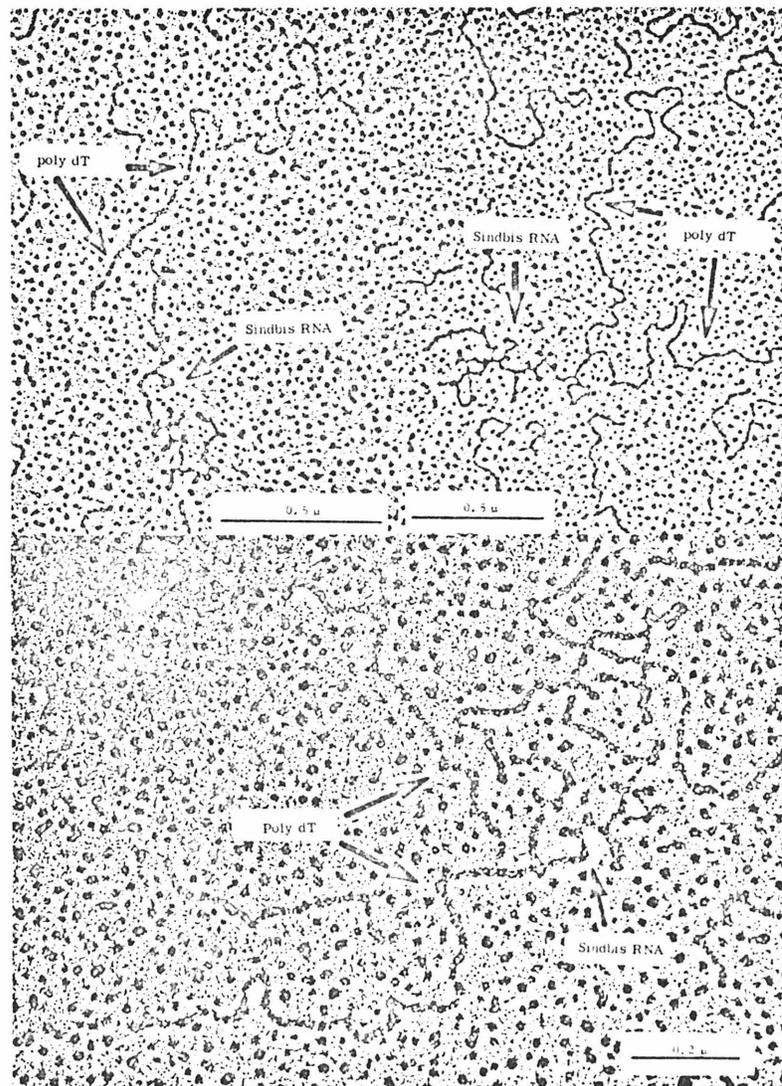


Figure 5

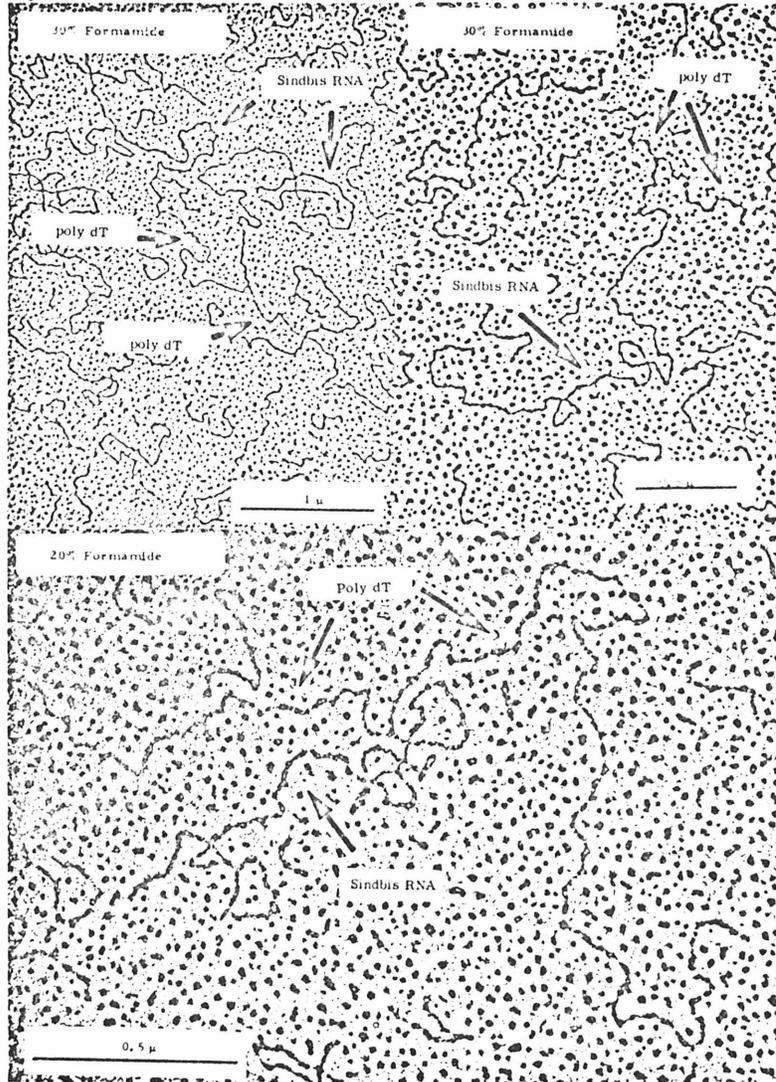


Figure 6

PROPOSITION 1

A New Method for Isolation of λ Trans-
ducing Phages Carrying Any Part of
E. coli Chromosome

Special transducing phages carrying a segment of bacterial DNA are useful for analyzing the genetic, physical and biochemical properties of bacterial genes. This is because by vegetative replication of the transducing phage, a large number of copies of a particular gene and gene products can be obtained for analysis.

Methods for constructing λ or $\phi 80$ transducing phages capable of carrying many parts of the E. coli chromosome have been established (1,2,3). These methods involve a transposition of the bacterial gene to be transduced to a site near the prophage or vice versa via an F-prime factor, or selecting lysogens with λ integrated at unusual sites. Several selection processes are necessary to achieve the final goal. Furthermore the number of F or λ integration sites are probably limited and there are preferred sites for such integration (3,4). Therefore these methods may not be useful for all regions of the E. coli chromosome.

I wish to propose a general method that is easy to apply. It is illustrated in Fig. 1. The basic idea is to add the power of random integration to a phage capable of specialized transduction. Mu phage is known to be able to insert into E. coli chromosome randomly.(5) However it has no transducing ability. The method is then based on

the construction of a λ -Mu hybrid phage carrying a small piece of Mu DNA in the λ genome (step 1). This can be achieved easily by having Mu inserted near λ (e.g. chlD), and then isolating a λ transducing phage carrying a small piece of Mu prophage. The small piece of Mu DNA is then used as a "guide" for integration of λ -Mu DNA into a prophage Mu by recombination between the homologous Mu DNA (step 2). This can be accomplished by infecting a bacterium that has a deletion in the chlD-bio region, and that has Mu inserted near gene A, and selecting for pgl⁺ transductants. Two kinds of structure for the inserted λ DNA are possible, depending on the relative polarity of the short Mu segment carried in λ and of Mu prophage in the bacterial chromosome. They are shown in Fig.2a and 2b respectively. Both structures are useful for constructing transductants. Since lysogens with Mu integrated at any part of the bacterial chromosome can be selected (except where the gene inactivated is essential), transducing phage carrying any part of the bacterial chromosome can be obtained.

The advantages of the method are : 1) A "universal" reagent (λ -Mu) can be obtained. Compared with F-prime factors used for transposition this is much more useful. 2) The number of selecting steps is small, so that the pro-

cedure is easier to carry out.

If λ -Mu carries the G loop of Mu (6), two kinds of λ -Mu transducing phages can exist, with the sequence of G loop inverted with respect to each other. Thus if these phages are used to transduce a bacterial gene, two orientation of the transduced gene relative to the λ DNA sequence can be obtained. (see Fig. 3) This can be used to isolate pure bacterial gene as is the case of lac gene (1).

STEP 1 λ lysogen

Mu

chID-Mu lysogeninduce λ

select λ pgl transducing phages. Test of transduction of pgl on a Mu lysogen with pgl deletion and the same strain without Mu prophage as a control.

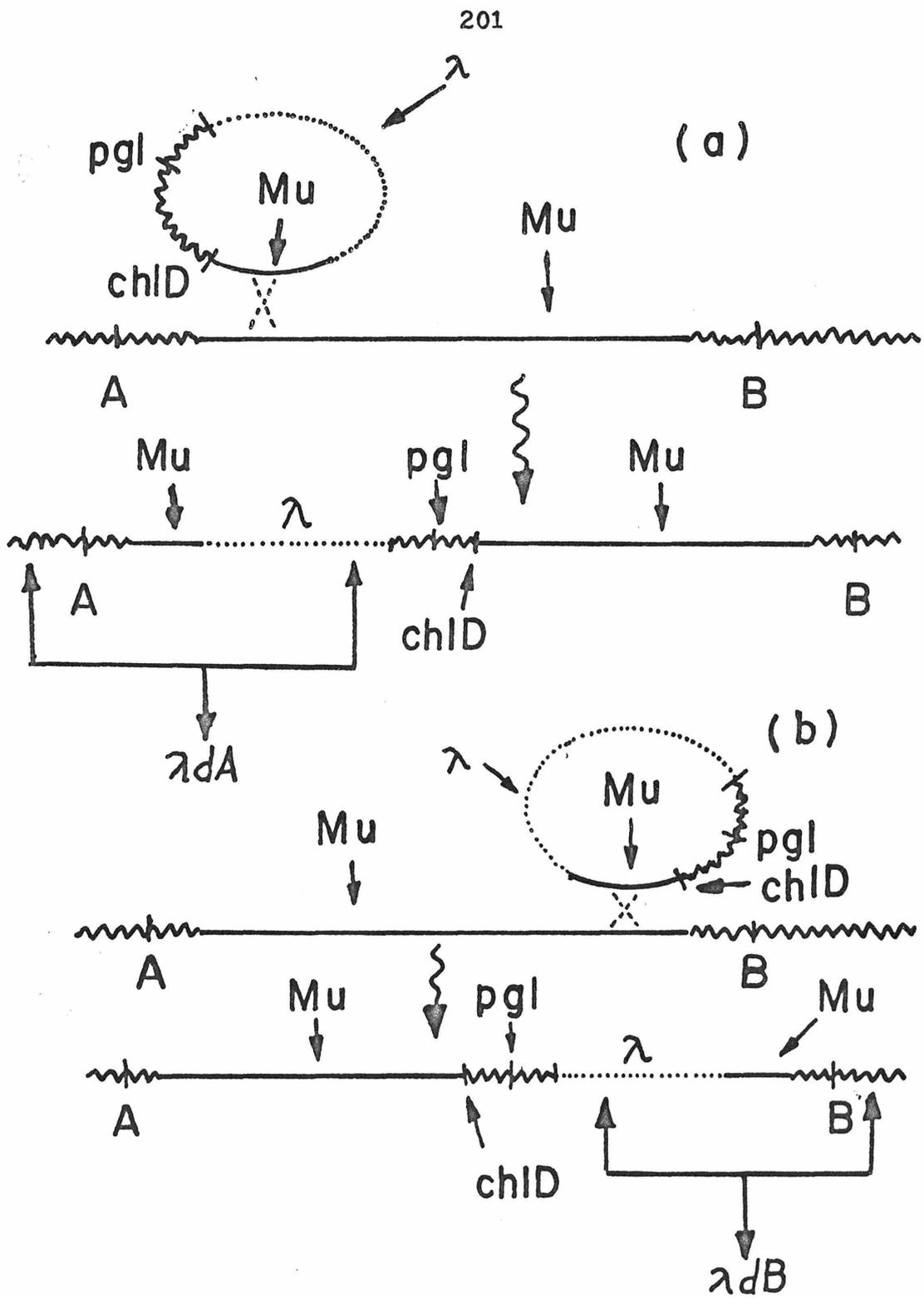


select λ -Mupgl which can transduce Mu lysogen at much higher frequency than the non-Mu lysogen.

STEP 2Mu lysogens (with chID-bio deletion) λ Mupglselect pgl⁺ transductants

induction

 λ transducing PhagesFIG. 1



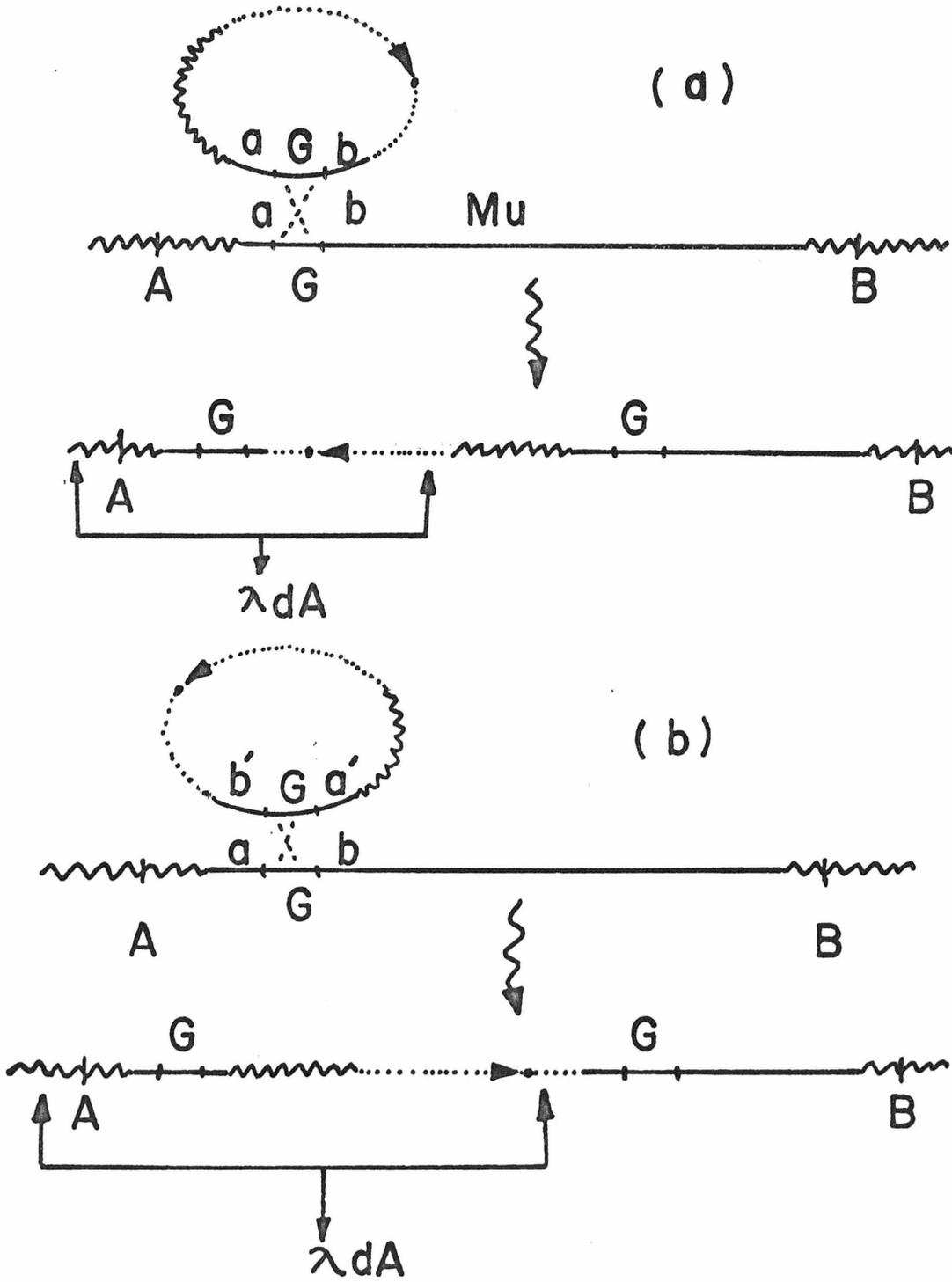


Fig. 3

REFERENCES

- 1) Shapiro, J., L. MacHattie, L. Eron, G. Ihler, K. Ippens, and J. Beckwith (1969). *Nature* 224, 768.
- 2) Gottesman, S., and J. Beckwith (1969). *J. Mol. Biol.* 44, 117.
- 3) Shimada, K., R. Weiberg and M. E. Gottesman (1972). *J. Mol. Biol.* 63, 483.
- 4) Falkow, S., and R. V. Citarella (1965). *J. Mol. Biol.* 12, 138.
- 5) Taylor, A. L. (1963). *Proc. Nat. Acad. Sci. USA*, 50, 1043.
- 6) Daniell, E., J. Abelson, J. S. Kim, and N. Davidson (1973) *Virology* 51, 237.

PROPOSITION 2

Application of the Diaminobenzidine
Technique for the in situ Localization
of Integrated Viral Genome and Single-
Copy Genes

The in situ hybridization technique has been used extensively in mapping DNA sequences in chromosomes of higher organisms. The method, however, suffers from some difficulties : 1) Only highly repeated genes can be analyzed. Many interesting problems such as the localization of the integrated viral genome in the transformed cell cannot be studied. 2) The autoradiograph technique used is difficult and requires a long period for a single experiment. The following proposition are designed to overcome these problems.

A. In situ localization of integrated SV40 and adenovirus genome in the transformed cell chromosome.

The scheme of the method is illustrated in Fig. 1. The basic idea is to convert a viral DNA into two functional units : a part with intact DNA sequence for hybridization with the sequence to be mapped, and a second part covered with an enzyme marker such as horseradish peroxidase for tracing the molecule. Linear viral DNA (SV40 is circular but can be converted into linear form by digesting with R1 restriction enzyme) is first digested with λ exonuclease to generate single strand ends. The single strand tails are then covered with single strand specific gene-32 protein of T⁴ phage(1) and

fixed with glutaraldehyde. Horseradish peroxidase is then covalently bound on gene-32 protein by using bifunctional reagent FNPS (p,p'-difluoro-m,m'-dinitrophenyl sulfone). This reagent has been used by Nakane and Pierce (2) to couple horseradish peroxidase and antibodies without destroying either enzymatic or immunologic activities.

The molecule thus constructed will contain a long intact DNA sequence capable of hybridization and a part with many peroxidase molecules which can be readily recognized in the light microscope by the diaminobenzidine technique (3). To localize the integrated SV40 genome in the chromosome of transformed cells, the enzyme coupled SV40 or adenovirus DNA is used to hybridize with the chromosome in situ and the chromosomes treated with diaminobenzidine, washed and then incubated with OsO_4 , and washed again. The sites of integrated SV40 genome will be recognized by the OsO_4 precipitate or the color of the enzyme reaction product in the light microscope.

B. Method for in situ localization of a single-copy gene in the chromosome

The idea is basically the same as in Sec. A. The detail of method is different because now mRNA is the starting material. It has been shown that all the mRNA (except that of histone) of eukaryotes have a stretch of polyA sequence at the 3' end. DNA complementary to the

RNA can be synthesized by using reverse transcriptase of oncornavirus and oligo dT as primer. This reaction is used to prepare a DNA complementary to the mRNA studied with an enzyme marker linked to the 5' end. The enzyme coupled oligo dT can be prepared by coupling with water soluble carbodiimide (4). The enzyme-oligo dT can then be used as primer in the synthesis of DNA complementary to the mRNA, as shown in Fig. 2. The enzyme covalently attached to the complementary DNA is used as a marker for in situ localization of single-copy genes.

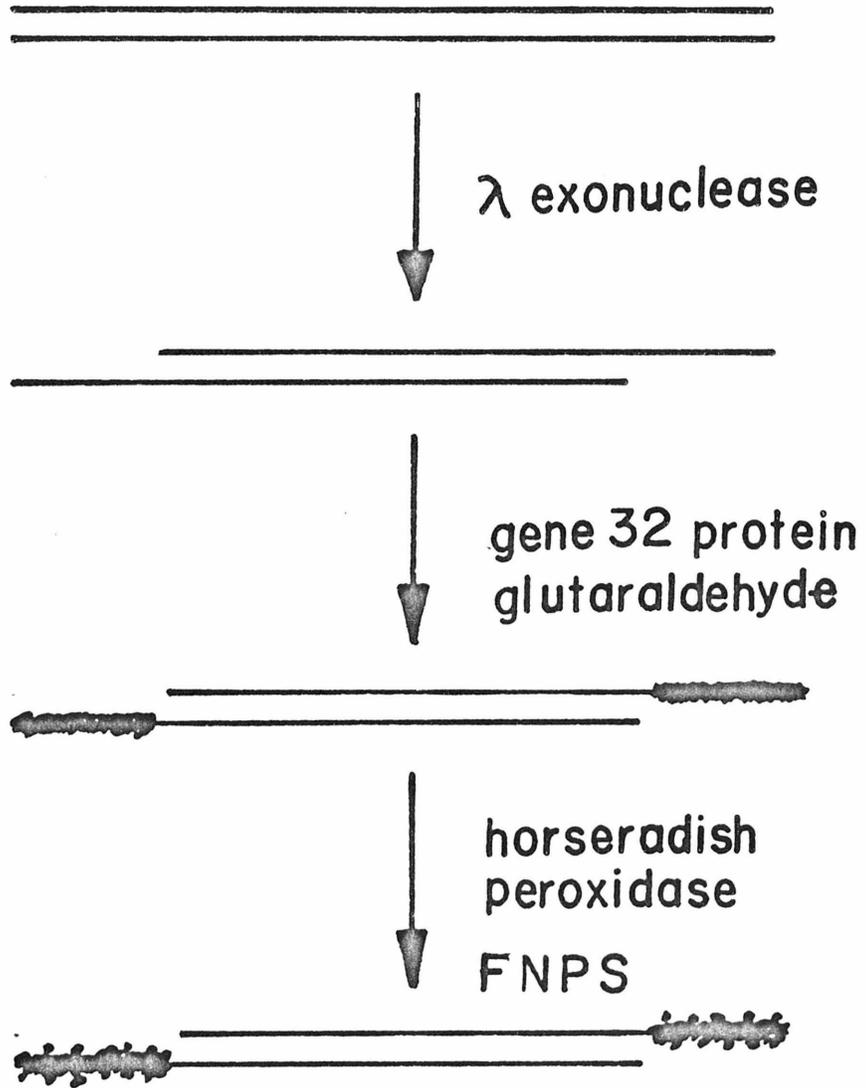


Fig. 1

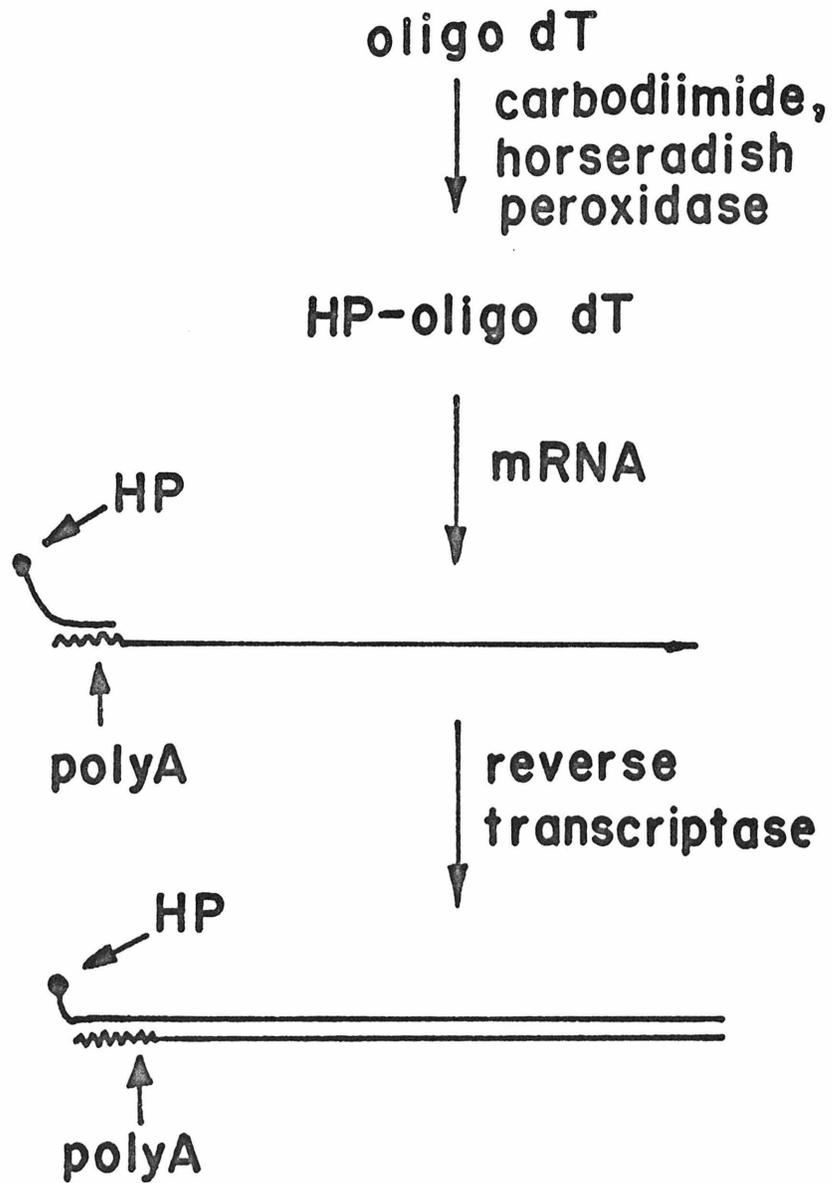


Fig. 2

REFERENCES

- 1) Delius, H., N. J. Mantell, and B. M. Alberts (1972)
J. Mol. Biol. 67, 341.
- 2) Nakane, P. K., and G. B. Pierce (1966) J. Histochem.
& Cytochem. 14, 929.
- 3) Shnitka, T. K., and A. M. Seligman (1971) Ann. Rev.
Biochem. 40, 375.
- 4) Gilham, P. T. (1968) Biochemistry 7, 2809.

PROPOSITION 3

Applications of an Inverse Immuno-
adsorption Technique in Isolating and
Analyzing Proteins

A protein can be purified from other components by adsorption to its antibodies fixed on a solid support (1). This useful technique, however, is not generally applicable in practice, because, in order to obtain antibodies a fairly large quantity of the purified protein has to be obtained first. It is proposed that this difficulty can be overcome by absorbing all other components except the molecule of interest, with an appropriate antibody preparation. This procedure avoids the need for purification of the protein of interest in order to obtain antibodies. The usefulness of this method is shown by the following examples.

I. Isolation of gene products in E. coli

Two bacteria strains A and B are used. B is derived from A by a single deletion in gene a for example. A whole cell protein extract of B is prepared and injected into a rabbit to induce antibodies. Since a whole cell protein extract is used, no purification steps are needed and the protein preparation can be obtained in large quantity. If necessary antibodies against B protein can be purified easily by adsorption with a column containing B protein bound on a solid support. The anti-B antibodies are then used to absorb the whole cell protein extracted from A. The gene product of gene a should be missing in B and therefore will not be absorbed. The column can be

repeatedly used by eluting out the proteins adsorbed. Thus large quantity of protein from E. coli. can be handled. The purification may not be complete but most of the impurities can be removed and therefore is useful for second step purification using other methods.

II. Isolation and analysis of gene products of viruses.

Gene products of viruses such T₄ are very complex. Demonstration of the absence of certain band in the gel electrophoretic pattern of the proteins extracted from a mutant virus is sometimes ambiguous. A minor band hidden in a major band may also be neglected. The virus proteins are in a large background of host proteins and therefore chemical analysis such as aminoacid composition is not possible. By preparing antibodies against the whole cell proteins extracted from a mutant phage infected cell, and use them to absorb whole cell protein extracted from a wild type phage infected cell (labelled with isotope), the gene products missing in the mutant phage can be readily recognized. As a control, protein extracted from wild type phage infected cells is absorbed by the antibodies and there no labelled protein should come out after repeated absorption. The phage gene product obtained is free from host cell protein background and can

be used for studying its chemical properties.

III. Study of nonhistone chromosomal protein patterns in differentiated cells

Nonhistone proteins have been inferred to be the specific control element in gene regulation (2). Thus nonhistone proteins isolated from cells of different tissues have been shown to have qualitative and quantitative differences (2). The analysis is however quite difficult due to the complexity of the nonhistone chromosomal proteins. The problem can be simplified by using the inverse immunoabsorption technique proposed. Antibodies against nonhistone proteins from cell type A are used to absorb labelled nonhistone protein from cell type B. The repeatedly absorbed protein is then run on the gel electrophoresis and proteins missing in cell A can be shown clearly. In the same manner nonhistone proteins missing in cell B can be analyzed.

REFERENCES

- 1) Cuatrecasas, P. (1972) Adv. in Enzymology 36, 29.
- 2) Markert, C. L. and Ursprung, H. (1971) " Developmental Genetics" Prentice-Hall Inc.

PROPOSITION 4

A Method for Eliminating Fluorescence
Background in Laser Raman Spectroscopy

Recently laser Raman spectroscopy has become a potential tool in studying biomolecules. This is because of the following advantages of Raman spectroscopy :

1) rich molecular information can be obtained, 2) the sample can be studied in aqueous solution at various temperatures, and 3) large molecules such as polynucleotides can be studied.

One of the difficulties in studying biomolecules using Raman scattering is that many Raman lines are obscured by the fluorescence of the sample. The problem is serious particularly in the low frequency region of the spectra and when the Raman line is weak.

It is the purpose of this proposition to devise a method to solve this problem. The idea is quite simple but no one seems to have thought about it at least to the most recent published paper I have read. The method is based on the shutter device invented by E. Gaviola in 1926 to measure the lifetime of fluorescence (1). A schematic illustration of the apparatus is shown in Fig. 1. (for explanation of the apparatus see the legend in Fig. 1) This device is based on the fact that fluorescence light is delayed relative to the scattered light by the decay time of fluorescence. (The usual decay time is of the order nanoseconds, see ref. 2). Therefore the fluorescence light entering the second Kerr cell is always out of phase

relative to the scattered light. The device can be set in such a condition that fluorescence light is prevented from entering the spectrometer.

The several advantages and applications using this method are the following:

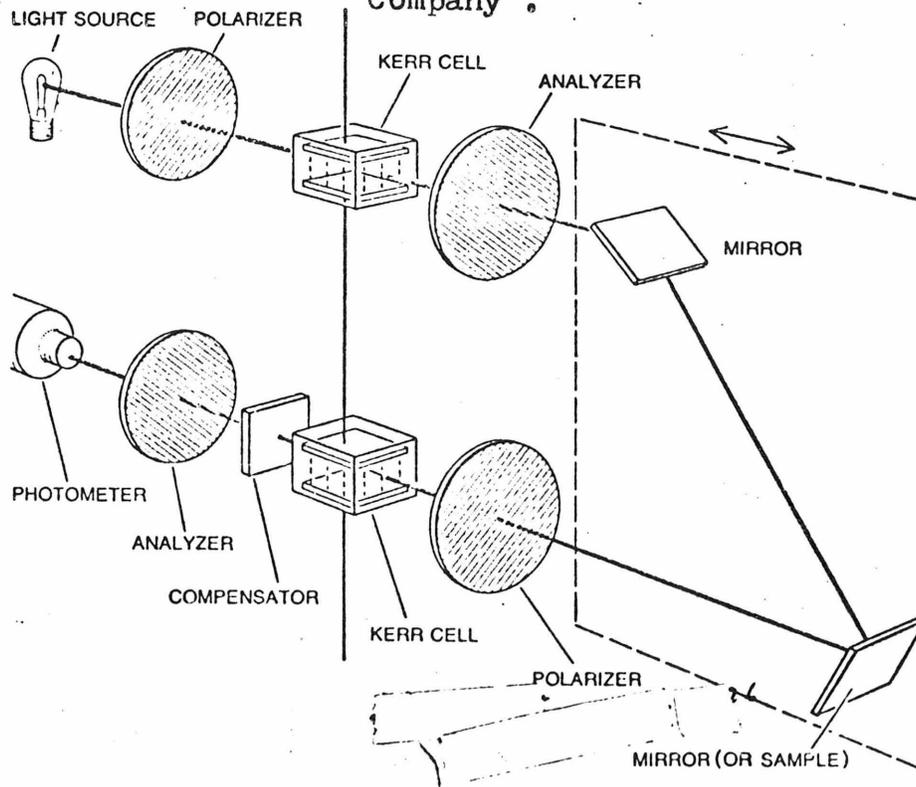
1) The painstaking steps to remove fluorescing impurities in the sample can be avoided.

2) The intrinsic fluorescence of proteins can be eliminated. This method when used together the iodine filter technique (3) to remove the strong Raleigh scattering can be used to obtain low frequency spectra of protein molecules. The Raman lines in this region ($10-200\text{ cm}^{-1}$) of the spectra are sensitive to the conformation change of protein molecules (4). The quality of the medium frequency region can also be improved. Thus the conformation sensitive C-S-S-C Raman lines (5) can be observed much easier.

3) Strong fluorescing sample such as flavoproteins can be studied.

4) Strong fluorescence background in Resonance Raman scattering experiments (6) can be eliminated.

The use of this figure is permitted by W. H. Freeman Company .



COMPLICATED SHUTTER SCHEME based on the principle of the direct-current Kerr gate was designed by Enrique Gaviola in 1926; an improved version of this scheme was later used to measure fluorescent lifetimes as short as a nanosecond (a billionth of a second) with an accuracy of a few percent. The apparatus consists of two Kerr cells hooked up in series with the voltage across them varying synchronously and sinusoidally at a characteristic frequency. Light from a light source is modulated electronically in the first Kerr cell. After being reflected by a system of mirrors the light passes through the second Kerr cell only if it arrives in phase with the modulation. If a fluorescent sample is placed at the position of one of the mirrors, the induced fluorescence from the sample will be delayed with respect to the frequency of the second Kerr cell by an amount that depends on the relaxation lifetime of the fluorescence. Phase differences can be adjusted by moving the part of the apparatus within the broken line, by varying the ellipticity of the polarization with a compensator or by changing the frequency. A plot of the intensity of the light received as a function of frequency enables one to derive fluorescent lifetime of the sample.

(from "Ultrafast Phenomena in Liquids and Solids"
Scientific American page 47, June, 1973.

Fig. 1

REFERENCES

- 1) Gaviola, E. (1926) Z. Physik. 35, 748.
- 2) Guilbault, G. G. (1973) "Practical Fluorescence, Theory Methods and Techniques" Marcel Dekker Inc.
- 3) Delvin, G. E., Davis, J. L., Chase, L. L., and Geschwind, S. (1971) Appl. Phys. Lett. 19, 138.
- 4) Peticolas, W. L. (1971) Biopolymers 10, 2223.
- 5) Lord, R. C. and Yu, N. (1970) J. Mol. Biol. 50, 509.
- 6) Tobin, M. C. (1971) "Laser Raman Spectroscopy" Wiley-Interscience Inc.

PROPOSITION 5

A New Method for Genetic Mapping in
Mammalian Cells

The first step in genetic mapping of mammalian cells is the assignment of a gene to a particular chromosome. This is usually done by chromosomal complementation through cell fusion between two cells of different species (1). This technique, which requires the analysis of chromosome karyotypes of a series of hybrid clones in order to determine a common chromosome responsible for a particular phenotype, is very time consuming, especially when the number of chromosomes is high. I propose that this difficulty can be largely overcome by taking the advantage of the recent finding that isolated metaphase chromosomes can be incorporated into a cell of different species (2,3,4).

The procedures are as follows: 1) select various mutants in Chinese Hamster cells (5). 2) isolate metaphase chromosomes of human or mouse cells for example. 3) incubate the chromosomes with the mutant cells in tissue culture and select for cells that are complemented through the incorporation of that chromosome which contains a normal genotype for the mutant trait. Since the probability of incorporation of chromosome is low, simultaneous incorporation of several chromosome is rare. Therefore the karyotypic analysis will be very simple. Probably only one or two cell clones will be needed to determine the chromosome carrying the normal genotype.

The most common problem for finding the complemented clones is the confusion due to reversion of the

original mutants. However since 1) reversion occurs at very low frequency (10^{-6} - 10^{-8}) 2) the reverted clones will contain gene products that are distinguishable from the complemented clones by either their immunologic or physical properties (e. g. electrophoretic mobility), it is possible to identify the complemented cell clones.

One interesting application of this technique is the mapping of integrated SV40 viral genome. Metaphase chromosomes isolated from mouse cell transformed by SV40 virus are mixed with Chinese Hamster cells which can also be transformed by SV40. Transformed clones of Chinese Hamster cells are then selected. Karyotype analysis of these cell lines will reveal the mouse chromosome that contain the oncogenic principle-the integrated SV40. Similar experiments in principle can be performed with cells transformed by RNA tumor viruses.

REFERENCES

- 1) Kao, F. T., and T. T. Puck. (1970) Nature 228, 329.
- 2) T. Sekiguchi, F. Sekiguchi and M. Yamada (1973)
Exptl. cell Res. 80, 223.
- 3) G. D. Burkholder and B. B. Mukherjee (1970). Exptl.
cell Res. 61, 413.
- 4) T. Ebina, I. Kamo, K. Takahashi, M. Homma and N. Ishida
(1970). Exptl. cell Res. 62, 384.
- 5) T. T. Puck. "The Mammalian Cell as a Microorganism"
Holden Day Inc. 1972.