# A STUDY OF THE BASE SEQUENCE ARRANGEMENTS IN DNA BY ELECTRON MICROSCOPY

Thesis by

Ronald Wayne Davis

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1970

(Submitted December 31, 1969)

ACKNOWLEDGMENTS

I want to express my appreciation to
Martha Simon, Sandy Parkinson, C. S. Lee,
Dave Clayton, Richard Hyman, and Rolf Knippers,
for collaborating;
the National Institutes of Health, for giving;
Norman Davidson, for guiding;
and
Janet Davis, for loving (and typing).

## Abstract

The study of base sequence arrangements in DNA molecules was accomplished by coupling electron microscopy (EM) with DNA hybridization. Modifications of the basic protein film technique were used. EM methods for examining both double-stranded and single-stranded DNA were used.

Many EM studies have a tendency to be rather descriptive in nature. However, it has been my aim to approach problems from a more quantitative aspect. The major problem in quantitative EM of DNA is length fluctuations. It was discovered that the standard deviation in the length of a homogeneous sample of DNA molecules is directly proportional to the square root of the length.

An EM method was developed for determining the size and location of deletion mutations and substitutions in the λ bacteriophage DNA. Heteroduplexes with one strand from the wild type and one strand from the mutant were formed by renaturation. The location of the nonhomology regions can be accurately mapped. A number of mutants were selected that are unable to integrate and by EM studies of the heteroduplexes it was discovered that there is a specific site in the λ DNA (0.574 from the left end) and a specific site in the E. coli DNA where the integration takes place. It was also shown that this specific site in λ DNA would not form stable base pairing with the

specific site in the E. coli DNA. This means that the two sites have a base sequence similarity of less than 10 base pairs. A physical gene map of the left arm of the λ DNA molecule was constructed by comparing the physical and genetic location of a large number of mutations.

The DNA isolated from a defective lysogenic phage of E. coli 15 was studied. After induction there is a closed circular DNA in vivo. However, the DNA is packaged as a linear molecule in the phage head. The mature linear DNA is 7.5% longer than the in vivo closed circular DNA. This was found to be because the mature linear DNA is both terminally repetitious (TR) and circularly permuted. Circular molecules are obtained by denaturing and renaturing the linear molecules. These in vitro circular molecules have two single-stranded DNA branches, each equal in length to the TR. The amount of the TR was found to be slightly heterogeneous. Therefore, the amount of linear DNA that is packaged in the phage head is slightly variable and there is no mechanism for packaging exactly the same amount of DNA.

The mitochondria of human leukemic leukocytes contain covalently closed circular DNA molecules of twice the molecular weight found in mitochondria of normal leukocytes. Heteroduplexes containing one dimer length strand and two complementary monomer length strands were formed by renaturation. These heteroduplex DNA molecules appear as

figure eights. They were carefully examined for DNA substitutions or deletions but none were found. Therefore, the dimer molecule is composed of exactly two monomer molecules.

EM studies of the SV40 DNA showed that it is heterogeneous in length. From heteroduplex studies it was learned that this DNA is heterogeneous in base sequence as well. About half of the self-renatured molecules contained nonhomology regions. The size (up to $\frac{1}{2}$ the molecule) and the number of non-homology regions varied considerably. Unique classes of heteroduplex molecules could not be found and no reasonable conclusions could be given for the nature or origin of the heterogeneity.

The replicating intermediates in the in vivo synthesis of ØX174 DNA were studied. The single-stranded DNA is synthesized from a double-stranded template as a linear molecule. The replicating intermediate appears as a double-stranded circular molecule with a linear single-stranded branch attached to the circle. The linear branch was never found to be longer than the mature circular DNA and is presumably released after one round of replication on the template and then cyclized by an as yet undetermined means.

RNA was synthesized on $T_7$ DNA in vitro. The resulting RNA-RNA polymerase-DNA complex was visualized in the EM. By synchronizing initiation and mapping the location

of the RNA at various times it was discovered that synthesis is always initiated close to one unique end, and thus, presumably at a unique site. At one min there is one RNA per DNA. At high RNA polymerase concentrations, closely spaced sequential initiations occur from the one active initiation site. The rate of propagation was measured as about 45 bases/sec.

TABLE OF CONTENTS


| CHAPTER | TITLE | PAGE |
|---|---|---|
| | A STUDY OF THE BASE SEQUENCE ARRANGEMENTS IN DNA BY ELECTRON MICROSCOPY | |
| 1 | Special Methods<br>A. Electron Microscope Heteroduplex Methods for Mapping Regions of Base Sequence Homology in Nucleic Acids<br>B. Separation of Lambda DNA Strands | 1<br>30 |
| 2 | Electron-Microscopic Visualization of Deletion Mutations | 36 |
| 3 | A Physical Study of the Integration of Lambda DNA into *E. coli* DNA | 45 |
| 4 | A Physical Gene Map of the Left Arm of the Lambda Chromosome | 108 |
| 5 | A Physical Study by Electron Microscopy of the Terminally Repetitious, Circularly Permuted DNA from the Phage Particles of *E. coli* 15 | 118 |
| 6 | Homology and Structural Relationships Between the Dimeric and Monomeric Circular Forms of Mitochondrial DNA from Human Leukemic Leukocytes | 187 |
| 7 | Base Sequence Heterogeneity in SV40 DNA | 253 |
| 8 | Single-Strand ØX174 DNA Synthesis: Electron Microscopy of the Replicating Intermediates | 270 |
| 9 | The Initiation Site for *In Vivo* RNA Synthesis on $T_7$ DNA | 289 |

# CHAPTER I

## A. Electron Miscrope Heteroduplex Methods for Mapping Regions of Base Sequence Homology in Nucleic Acids.

Ronald W. Davis, Martha Simon, and Norman Davidson

The following paper is in press in

Methods in Enzymology.

# Electron Microscope Heteroduplex Methods for Mapping Regions of Base Sequence Homology in Nucleic Acids

By Ronald W. Davis, Martha Simon, and Norman Davidson

It is, at present, either difficult or, more frequently, not possible to determine the base sequence of a high molecular weight nucleic acid by any direct method. However, the question of whether two nucleic acid strands have complementary sequences can be answered by hybridization experiments. We describe here a new type of hybridization experiment in which two partially complementary, partially noncomplementary strands are allowed to renature and the reaction product examined in the electron microscope. The regions of homology and of nonhomology can be mapped, because single- and double-stranded nucleic acids are recognizably different in suitable electron microscope preparations.[1,2]

If a mixture of two related double-stranded DNA's, AA' and BB', is denatured and renatured, the reaction mixture will consist of: the homoduplexes AA' and BB', the heteroduplexes AB' and A'B, and unrenatured single strands. The heteroduplexes are the molecules of interest here.

The heteroduplex method will, we anticipate, be broadly applicable; however, most of the studies so far have been on bacteriophage DNA's. The techniques described below are to some extent specialized for this case.

Techniques

1. Renaturation. When examining heteroduplexes in the electron microscope, a major cause of uncertainty and complications in interpretation is the presence of fragments due to single-strand breaks. When two intact, closely related single strands renature, the resulting molecule is largely two-stranded and does not renature further. When fragmented strands renature, the product is more likely to have a sufficient length of exposed single strands so that further renaturation to give branched and generally misleading aggregates occurs. Length measurements and quantitative mapping are very difficult if one does not have intact strands.

The procedures for DNA preparation, denaturation, and renaturation should therefore be designed to minimize the number of single-strand breaks in the DNA. We find, for example, that lambdoid phages grown on E. coli strains C600 and W3110 have satisfactorily intact DNA (about one single-strand break per 10 strands); phages grown on strain 594 have a large number of single-strand breaks. Phage are finally purified by banding in CsCl and are stored in 4 M CsCl, 0.01 M $Mg^{++}$, 0.01 M Tris, pH 8. The DNA is best stored in the phage. Native DNA can be prepared by heat and/or chelate shocking of the phage, or by treatment with urea or $NaClO_4$. The phage ghosts or phage protein present do not interfere with electron microscopy. Denatured DNA is prepared directly from the phage by simultaneous lysis and denaturation with alkali. For heteroduplex experiments

starting with closed circular DNA's, it is desirable to introduce only one single-strand break per duplex. This can be done by first separating the closed circular molecules and then lightly nicking, either with DNase[3] or with visible light in the presence of ethidium.[4] Despite all precautions, one never has perfectly intact DNA. Therefore, to minimize problems due to higher aggregates, we recommend underrenaturing the DNA. Renaturation of approximately 50% is optimal. Unrenatured single strands of DNA are easily recognized and can be ignored.

For most cases, there is no advantage in separating complementary strands[2] and we recommend avoiding the additional manipulations involved.

Renaturation in aqueous solution requires a high salt concentration and heating to about ($T_m$ - 30°C). However, the heating causes single-strand breaks. A short heating procedure (20 sec) at a high DNA concentration ($A_{260} \approx 1.0$) has been successfully used.[1]

Renaturation in a formamide solvent is a better procedure.[2] One may use moderately dilute DNA solutions ($A_{260} \approx 0.1$), room temperature or less, and longer times. Some data on the effects of temperature and ionic medium on the thermodynamics and kinetics of renaturation in formamide are available.[5,6] The effects of complexity, molecular weight, temperature, and ionic strength can also be estimated from studies of renaturation kinetics in aqueous solution.[7,8] The detailed description below applies to lambdoid phage DNA's.

The renaturation temperature (25° C) is about 30° below $T_m$, which gives an optimal rate; at lower temperatures it is much slower. Formamide solutions should be well buffered since they tend to drift toward a pH of about 5. Denaturation and complete strand dissociation are achieved, for example, by allowing a solution containing $5 \times 10^{10}$ particles (2.5 μg DNA) of each of the two phages in 0.5 ml 0.10 M NaOH, 0.02 M EDTA to stand for 10 min at room temperature. One then adds 50 μl of 1.8 M Tris·HCl, 0.2 M Tris·OH, and 0.5 ml of formamide. The pH usually reads 8.5. About 50% renaturation occurs in one to two hours, and the pH is then usually between 7.5 and 8.0. Renaturation is stopped by dialysis against 0.01 M Tris, 0.001 M EDTA, pH 8.5, at 4° C. Alternatively, the formamide solution is cooled to 0° C to slow down the renaturation reaction and immediately mounted for electron microscopy.

In an alternative ultramicro procedure, a mixture of two phages containing 0.01 μg DNA in 10 μl of solution is placed in 1/4" dialysis tubing, with knots tied about 1 cm apart. Each dialysis generally takes about one hour. Lambdoid phages are lysed by dialysis against 0.001 M EDTA (pH 8.2); or 6 M $NaClO_4$; or 8 M urea, 0.1 M NaCl, 0.01 M EDTA, pH 8.2. The $NaClO_4$ or urea are removed by dialysis against 0.01 M EDTA, pH 8.2. Denaturation is caused by dialysis against 95% formamide, 0.01 M EDTA, pH 8.2. Renaturation is achieved by dialysis against 50% formamide, 0.1 M Tris, 0.01 M EDTA, pH 8.2 for 30 min. The sample is removed and adjusted to

the appropriate reagent concentrations for electron microscopy in 30 μl of solution. A minor disadvantage of this method is that it is difficult to achieve precise control of the degree of renaturation.

2. Mounting DNA for Electron Microscopy. We will describe two variations of the basic protein film technique.[9,10] In the aqueous technique double-stranded DNA appears, as usual, as a gently curved filament (worm-like coil). Single-stranded DNA is condensed into bushes because of random base-base interactions. The technique is useful for measuring double-strand lengths, measuring the positions of single-strand regions along the double strand, and distinguishing positively between single- and double-stranded DNA.

In the formamide technique, the DNA is mounted at a formamide and salt concentration such that double-stranded DNA is stable, but the random base interactions in single-strand DNA are melted out and single-strand DNA also appears as a curved filament which is distinguishable from double-stranded DNA by being thinner and somewhat more kinky.[2,11] Of the several denaturing agents that we have tried--formamide, dimethyl sulfoxide, N-methyl formamide, purine, caffeine, methylmercuric hydroxide, urea, formaldehyde, sodium perchlorate, heat, and pH 12--we prefer formamide, the use of which for electron microscopy of DNA was first described by Westmoreland, et al.[2] This technique enables one to observe the topology at the connections between single- and double-stranded regions, thus distinguishing, for example, between deletions and substitutions; it also makes it possible to measure single-strand lengths.

Our own experience is mainly with cytochrome c. Several other authors report that certain globular, small, basic proteins, including cytochrome c, lysozyme, chymotrypsin, and trypsin are suitable for forming the protein film which binds the DNA.[10] In our hands, histones give poor results. The structure visualized in the electron microscope is a collapsed column of basic protein around the DNA which is 80 to 150 Å thick; it can be shadowed, or stained with uranyl salts.[12, 13] We pick up the protein-DNA mixed film on grids covered with a freshly spread (3-48 hours) Parlodion film from a 3.5% w/v solution in amyl acetate. These supporting films are quite strong and generally withstand the electron beam at crossover.

The contrast with either staining or shadowing depends on the amount of protein collapsed around the DNA, which itself depends on many factors. Two major variables are the ionic strengths of the spreading solution and the hypophase. Increasing the ammonium acetate concentration increases contrast, but if the concentration in the hypophase is above 0.5 M, little or no DNA sticks to the film. Furthermore, there is an increasing tendency to form lateral aggregates when the hypophase concentration is raised above 0.4 M. Flowerlike aggregates sometimes form below 0.2 M $NH_4Ac$ in the hypophase. Therefore, in the aqueous technique, we use 0.5 M and 0.25 M $NH_4Ac$ for spreading solution and hypophase, respectively.

In the formamide technique, there is less tangling of single-strand

DNA at low ionic strengths. Most denaturing agents reduce the contrast, formamide raises it. Increasing the formamide concentration increases the contrast; 85% formamide for double-strand DNA and up to 95% formamide for single-strand DNA have been used with good results. The distinction between double- and single-strand DNA is greater when there is only a small amount of protein around the DNA, and thus at a low ionic strength and with just enough formamide to extend the single strands. Shadowing is better than staining for making this distinction. We use 0.1 M $NH_4Ac$, 0.01 M Tris, 40% formamide, pH 8, for the spreading solution, and 0.01 M Tris, 10% formamide for the hypophase.

In our experience, the following reagents and conditions generally cause deleterious effects in the basic protein technique: (a) acidic conditions [depurination (pH < 7) and aggregation (pH < 6)]; (b) basic conditions (pH >10) (poor contrast and aggregation); (c) detergents (aggregation); (d) nitrate salts (poor contrast); (e) many denaturing agents (poor contrast). As a salt we recommend $NH_4Ac$; as a buffer, Tris plus EDTA.

In order to measure the position of a particular feature, say, a deletion or a substitution, from the end of a heteroduplex, the molecule must be fairly well straightened out with no or only a few crossover points, and must be well contrasted. These features are very dependent upon what appear to be minor variations in technique: the amount

of talc used, the way the hyperphase is allowed to run down the glass ramp, whether the supporting film is carbon or parlodion, the way the protein film is picked up, the position with respect to the ramp from which the film is picked up, and the length of time the film is allowed to stand, as well as the obviously important features of the composition of spreading solution and hypophase. The length of a single-stranded DNA varies markedly with minor variations in conditions; the length of a double-stranded DNA varies over a smaller fractional range. Conceivably, the use of a film balance would give more reproducible and definable conditions. However, specimens picked up from different positions of a single cytochrome film have different characteristics; we therefore suspect that the single parameter of film pressure is not sufficient to define conditions in the film.

The only good way to communicate details of technique is visually (through a film strip, for example). We shall do the best we can in writing, and the reader must expect to experiment and practice before achieving a satisfactory technique.

2A. Aqueous Technique. The concentrations in the spreading solution are as follows: 0.5 μg/ml of DNA, 0.1 mg/ml of cytochrome c, 0.5 M $NH_4Ac$, and 0.001 M EDTA (pH 7.5). The hypophase in a 90 millimeter square plastic Petri dish is 0.25 M ammonium acetate (pH 7.5). A little talc can be dusted on the hypophase in order to visualize and compress the film. The glass slide serving as a ramp should be

rinsed with 0.25 M $NH_4Ac$ and allowed to drain dry before applying the spreading solution. A quantity of 50 µl of the spreading solution is spread via a teflon tubing-Hamilton syringe combination back and forth across the slide starting about 1 cm above the surface of the hypophase and ending as the spread solution contacts the hypophase. The solution runs down the slide and the protein film spreads, pushing the talc away. For best results the film is picked up at once onto the grid at a distance of 1 grid width from the slide-solution boundary. The grid is dipped for 30 sec into the uranyl stain solution, then rinsed for 10 sec in isopentane. The stain solution contains a concentration of uranyl acetate or uranyl chloride of $5 \times 10^{-5}$ M in 90% ethanol and is freshly prepared (within 1 hr) by diluting a stock solution which is $5 \times 10^{-2}$ to $5 \times 10^{-3}$ M in the uranyl salt, 0.05 M in HCl, in 96% ethanol, and which is always stored in the dark. The more concentrated stock solution, when diluted, generally gives a more highly contrasted DNA, but also more uranium oxide precipitate in the background. The contrast can be increased by rotary-shadowing with heavy metal but this is not recommended when single strands are present. Excellent contrast can be obtained with the aqueous technique by staining alone.

The grain size for stained DNA is very small and single-strand bushes have the same stain density as duplex DNA. As a result single-strand bushes are easily identified when stained, but are often obscured by shadowing.

2B. Formamide Technique. For most purposes the spreading solution should contain 0.5 μg/ml of DNA, 0.1 mg/ml of cytochrome c, 0.1 M Tris and 0.01 M EDTA (pH 8.5) in 40% formamide. The concentration of formamide needed to denature DNA at this salt concentration and in the presence of cytochrome c is about 90%. We have used 50 to 70% formamide when it is desired to minimize duplex formation due to relatively short sections of complementary base pairs. The hypophase is 10% formamide, 0.01 M Tris (pH 8.5) and should be made up shortly (5 min or less) before use since it may become acidic relatively quickly. The spreading solution should be used within an hour or two. A clean glass slide, pretreated with $NH_4Ac$ solution and allowed to drain dry, is put into the Petri dish containing the hypophase. Little or no talc should be used. Fifty μl of the spreading solution is slowly applied from the teflon tubing at the slide-solution boundary. The film is allowed to stand for 1 min before being picked up. This seems to help in spreading out the single-strand DNA. The film is picked up near the slide-solution boundary and is again stained and rinsed. Some increase in contrast is obtained if the film is slightly compressed with a teflon bar before picking it up. In general, with the formamide technique, when the DNA is only stained the contrast is poor and the DNA can only be easily visualized by dark field electron microscopy. The contrast, however, is quite adequate if the DNA is also shadowed. When

staining, the more dilute stock solution of uranyl acetate should be used to reduce the amount of uranyl oxide background. Indeed, most of the contrast results from the shadowing, and the staining procedure can be omitted.

3. Measurement Procedures. Only molecules with neither intra- nor intermolecular overlaps should be measured in order to avoid ambiguities. Whenever possible, any DNA's added for calibrations should appear in the same electron micrograph as the heteroduplex of interest. Our measurements are taken on 35 mm film at magnifications from 1,000 to 10,000 X with a Philips EM 300 electron microscope using a 50 μ objective aperture and 60 kv accelerating voltage. The resulting negatives are enlarged 10 to 50 X on a Nikon shadowgraph and are traced on tracing paper. Lengths are then measured with a Keuffel and Esser 620 300 map measurer. The enlargement should be at such a ratio that the map measuring errors are not significant. In general, from 20 to 100 molecules should be measured depending on the accuracy desired.

## Results and Interpretation

1. Examples. The heteroduplex technique has been used for mapping the position and length of deletion mutations in λDNA, for studying the position and length of pieces of nonhomologous DNA which have replaced a piece of the original DNA (for example, in transducing phages such as λdg's), for studying the regions of sequence homology

and nonhomology in the DNA molecules of related phages, and for studying the relationship between the sequences occurring in the circular dimers and monomers of mitochondrial DNA. Examples of these four applications are shown in Fig. 1.

Heteroduplex techniques can be used to demonstrate that certain bacteriophages have linear DNA molecules which are circularly permuted and terminally repetitious. When such a DNA is denatured and renatured, circular duplex molecules with protruding single-strand branches are formed.[14] If the renatured molecules are observed under formamide conditions, the length of the terminal repetition and its length heterogeneity can both be measured.[15]

A micrograph such as that of the 434/λ heteroduplex in Fig. 1 is interpreted to mean that in the regions which are double-stranded, 434 DNA and λDNA have the "same" base sequences and in the regions where single-strand loops occur the base sequences are "nonhomologous". It must be emphasized that at present we do not know how perfectly complementary two sequences must be to give a double-stranded structure and how little complementarity there must be before one sees loops and concludes that the base sequences are nonhomologous.

Although the regions of homology and nonhomology in the lambdoid phage heteroduplexes are generally no smaller than 1,000 base pairs, we believe that regions as small as 50 base pairs can be detected.

2. Length Measurements. Two important points will be discussed in this section. The first is that since the length of a DNA molecule, single- or double-stranded, depends on the conditions under which it is mounted for electron microscopy, absolute length measurements as calibrated with a diffraction grating are not very significant. However, by measuring length relative to appropriate DNA standards, quantitatively significant results can be obtained. The second point is that for a homogeneous DNA, in which all molecules have the same number of nucleotides in the same base sequence, it is an inherent property of the basic protein film method that the length of the DNA molecule fluctuates around a quantitatively significant mean value with a quantitatively reproducible standard deviation. Hence, by using length measurements relative to a homogeneous standard both the mean value for the length of the unknown and the standard deviation of its length about the mean are reproducible and significant quantities.

The absolute length (in microns) of both double- and single-stranded DNA is dependent upon the conditions of mounting the DNA for electron microscopy. For example, the absolute length of both double- and single-stranded DNA decreases with increasing ionic strength.[16] Single-stranded DNA lengths are markedly sensitive to the formamide concentration. Below 30% formamide (spreading solution) the single-strand lengths increase with increasing concentration of formamide; above this value they decrease. Both single- and double-strand lengths, but especially the former, are subject to

uncontrollable variations from grid to grid even when the grids are mounted under seemingly identical conditions. The relevant variables, we believe, are concerned with the film pressure and density and the amount of cytochrome collapsed around the DNA, but it is not known how to control these variables. Instrumental magnification errors and variations in focal conditions may also contribute, although we are inclined to believe that these are less significant.

An effective, quantitatively significant, way of obtaining molecular lengths and thus molecular weights of segments or of a whole molecule is to mount and photograph a standard DNA on the same grid as the unknown. (The calculation of molecular weights by comparative contour length measurements makes the as yet untested assumption that the length per base pair is independent of base composition and sequence.) Two methods of using such internal calibrations will be mentioned. In the simpler and more generally applicable method one adds a single- and/or a double-stranded DNA, each of known length, and each clearly distinguishable from the other standard and from the unknown. For a single-stranded DNA, we use $\varphi$X 174, which can be recognized by its circular form. For a double-stranded DNA we either use the nicked double-stranded replicative form of $\varphi$X 174 RFII DNA, or a $\lambda$ phage DNA of known length. The single- and double-stranded $\varphi$X's can be distinguished by their appearance.

We take $1.7 \times 10^6$ as the molecular weight of $\varphi$X 174 DNA and thus $3.4 \times 10^6$ for the $\varphi$X RF.[17] We find that the ratio of length of

double-stranded λ DNA to φX RF DNA is 9.1(± 0.1) so that the molecular weight of the λDNA is calculated to be $31.0(\pm 0.3) \times 10^6$ which is within the range of values reported by other investigators by several methods.[18-21]

The second method of length calibration involves making a heteroduplex of the unknown DNA with a known DNA that contains a particular substitution or deletion at a known position and of known length.[22] This method also serves to orient the position of the deletion with respect to the right and left ends of the molecule. Suppose, for example, that it is desired to map the positions of the deleted region in the deletion mutant $\lambda_{b522}$. Its position can be accurately mapped by forming, for example, the heteroduplex $\lambda_{b522}/\lambda_{b536}$. The b536 is a small deletion in the left arm of λ with its left and right termini previously mapped at $\underline{x}_1^{536} = 0.426$ and $\underline{x}_2^{536} = 0.463$ in units of the length of λDNA.

As illustrated in Fig. 2 and shown in Fig. 1a, the heteroduplex $\lambda_{b\,522}/\lambda_{b\,536}$ has two bushes. The micrograph in Fig. 1 shows that the b522 deletion is to the right of the b536 deletion. The distances (in arbitrary units) from the left end of the molecule to the b536 bush, from the b536 bush to the b532 bush, and from the b522 bush to the right end are measured on the same molecule as $\underline{X}_L^{536}$, $\underline{X}_{536}^{522}$, $\underline{X}_{522}^{R}$, respectively. Then the fractional coordinates of the left and right termini of the b522 region are

$$\underline{x}_1^{522} = \underline{x}_2^{536} + (\underline{X}_{536}^{522}/\underline{X}_L^{536})\underline{x}_1^{536}$$

$$\underline{x}_2^{522} = 1 - (\underline{X}_{522}^{R}/\underline{X}_L^{536})\underline{x}_1^{536}$$

The numerical values are $\underline{X}_{536}^{522}/\underline{X}_L^{536} = 0.261$ and $\underline{X}_{522}^{R}/\underline{X}_L^{536} = 0.838$, so that $\underline{x}_1^{522} = 0.574$ and $\underline{x}_2^{522} = 0.643$.

We find empirically that there is an intrinsic fluctuation in length of a given homogeneous double-strand DNA so that the variance ($\sigma^2$) around the mean is proportional to the mean length $\langle \underline{L} \rangle$ of the DNA. The experimental data are shown in Fig. 3. Thus, we find for the standard deviation of length, $\sigma_D$, the equation

$$\sigma_D = \underline{D}(\underline{L}_D)^{1/2} \qquad (1)$$

The same length fluctuations are found for double-strand segments of a heteroduplex bounded by two bushes (or single-strand loops when mounted by the formamide technique) or a bush and an end. The length dependence of $\sigma_D$ holds between molecular weights of $3 \times 10^4$ and $3 \times 10^7$. If lengths are measured in units of the length of $\varphi$X 174 RF molecules, the empirical constant $\underline{D}$ has the value 0.037.

In eq. 1, subscript $\underline{D}$ denotes double-strand DNA. A similar plot of the standard deviation of length, $\sigma_S$, against the square root of the mean length, $\langle \underline{L}_S \rangle^{1/2}$, for single-strand DNA is also shown in Fig. 3. The data appear to conform approximately to the equation

$$\sigma_S = \underline{S}\langle \underline{L}_S \rangle^{1/2} \qquad (2)$$

However, many of the data points are for loops in heteroduplexes, for which there may be additional causes of variability. Additional samples of homogeneous single-strand DNA's spanning a range of mean lengths need to be measured to test the validity of eq. 2. The approximate value of the constant $\underline{S}$ is 0.049, if single-strand lengths are measured in $\varphi$X-174 units.

The empirical fact that equation (1) above applies may be interpreted to mean that there is some fundamental segment length for DNA molecules mounted in basic protein films, and observed in the electron microscope. The contour length of a molecule is the sum of the segment lengths. There is an intrinsic variability in the length of a segment, and each segment length is a stochastically independent variable.

Given, eq. 1 the principles of statistics assert that if sets of measurements are made on samples of $\underline{n}$ molecules, the mean lengths, $\langle \underline{L} \rangle$, of the samples should be normally distributed with a standard deviation of $\sigma_D/\underline{n}^{1/2}$ where $\sigma_D$ is calculated from eq. 1. Therefore, if contour lengths, $\underline{L}_a$ and $\underline{L}_c$, are measured on a sample of $\underline{n}_a$ and $\underline{n}_c$ molecules of type A and calibrating molecules, type C, of known length $\langle \underline{L}_c \rangle$, respectively, then the standard sample error, $\epsilon_a$, in the measurement of $\underline{L}_a$ is given by

$$\frac{\epsilon_a}{\langle \underline{L}_a \rangle} = \underline{D}\left( \frac{1}{\underline{n}_a \langle \underline{L}_a \rangle} + \frac{1}{\underline{n}_c \langle \underline{L}_c \rangle} \right)^{1/2} \qquad (3)$$

Thus it is advantageous to have a long DNA as the standard.

For example, for a double-strand DNA of length 9.0 φX RF units, the standard deviation in length for an infinite sample is predicted to be 0.11 φX lengths or 0.012 λ lengths. If an unknown, of approximately λ length (but sufficiently different to be distinguished on the micrographs) is measured relative to λ, and 25 molecules of each type are measured, the predicted fractional standard sample error, $\epsilon_a/\langle \underline{L}_a \rangle$, is 0.004. These calculations are confirmed in practice.[22]

We denote the variance of length of a sample of size $\underline{n}$ by $\underline{s}^2$. The principles of statistics assert that the quantity $\underline{n}\,\underline{s}^2/\sigma^2$ is a stochastic variable with a $\chi^2$ distribution for $\underline{n}$ -1 degrees of freedom. For reasonably large values of $\underline{n}$, the values of $\underline{s}$ in a population of samples are approximately normally distributed around $\sigma$, with a variance of $\sigma^2/2(\underline{n} - 1)$.

For the sample above, the standard deviation in length for the 25 molecules should lie with 95% probability within the interval 0.70 $\sigma$ to 1.25 $\sigma$, where $\sigma$ = 0.012 λ lengths. In particular, if the measured $\underline{s}$ is larger than 1.3 $\sigma$, some other cause of length heterogeneity, such as the presence of broken molecules or a true heterogeneity in the number of base pairs per molecule, should be suspected.

3. Assigning Single-Strand Loops. With a complicated heteroduplex pattern, such as that shown for the 434/λ heteroduplex in Fig. 1, the assignment of single-strand regions to either 434 or λ presents a problem. The following techniques are helpful and frequently sufficient

for an unambiguous assignment.[24] The total double-strand length plus the length of the single strands assigned to λ, suitably standardized by the calibrating single- and double-stranded DNA's, should add up to the total double-strand length of λDNA. Similarly, the double-strand length plus the length of single-strand loops assigned to 434 should add up to the double-strand length of 434. Another very helpful trick is illustrated by the following example. There is a loop in the middle of the 434/λ heteroduplex (Fig. 1d) in which the two single strands have lengths of 16% and 10%, respectively. Which of these strands belongs to λ? A heteroduplex, $434/\lambda_{b540}$ is made. The deletion mutant, $\lambda_{b540}$, has a deletion of 12% of λ length in the region of this particular 434/λ loop. In this heteroduplex, there is a loop with single strands of length 4% and 10% of λ. Therefore, the 16% strand of the 434/λ heteroduplex is assigned to λ and the 10% strand to 434.

4. Demonstrating Sequence Homology. The following problem sometimes arises. A mixture of 2 DNA's, AA' and BB', is denatured, renatured, and examined in the electron microscope. Only perfect double-strand molecules are seen. Either AB' and BA' are perfectly homologous (except for point mutations or other features that are not recognized as indicating nonhomology in the electron microscope) or they are completely nonhomologous and no heteroduplexes AB' have formed. The latter possibility can be tested in two ways.

(a) If the complementary strands, A and A', and B and B', can be physically separated, for example, by the poly-UG buoyant banding

procedure,[23] then pure A and pure B' are renatured together. Any double-stranded molecule seen has to be AB'.

(b) An alternative procedure[25, 15] which avoids the necessity of isolating separated strands is to shear one of the DNA's, say BB', to small pieces and renature a mixture of unsheared AA' and sheared BB'. If there is sequence homology between the sheared and the unsheared DNA's, heteroduplexes, each of which consists of a sheared strand and an unsheared strand will form. These heteroduplexes can be recognized as molecules with a short inner double-strand region and two single-strand ends, one or both of which is large (see next section). The DNA is trapped in the basic protein film from aqueous solution so that single-strand regions appear as bushes. The length of a single-strand piece can be approximately estimated from the perimeter of the bush.[1] The molecules of interest are the whole/fragment heteroduplexes. These are readily distinguished from whole/whole homoduplexes which have a long double-stranded region and fragment/fragment homoduplexes which have only small end bushes. The occurrence of whole/fragment homoduplexes due to the presence of single-strand breaks in the unsheared AA' preparation can only be evaluated by the control sample of renatured AA' alone.

5. The Excluded Volume Effect. This effect influences various heteroduplex experiments and its implications should be considered in any given experiment. Consider a single-strand DNA as a random coil

molecule. Random walk theory shows that a point near the topological end of the strand is, on the average, closer to the physical outside of the coil than is a point close to the topological center of the strand. Because of excluded volume effects, a point on the physical outside of the coil is more available for the initiation of a renaturation reaction with a homologous segment from another DNA strand. This effect predicts, in accordance with observations,[25, 26] that in the formation of whole/fragment heteroduplexes, as described in the preceding section, there is a greater probability of a fragment homologous to the topological outside of the whole strand renaturing than there is for a fragment homologous to the topological center, even though there are equal concentrations of the two kinds of fragments in the solution. Therefore, most of the whole/fragment heteroduplexes have unequal size end bushes. The excluded volume effect predicts also that in the renaturation of a population of circularly permuted DNA's there is a preference for mutual renaturation between complementary strands which have their beginning points fairly close to each other along the genome.[15]

Footnotes

[1]R. W. Davis and N. Davidson, Proc. Nat. Acad. Sci., U.S., 60, 243 (1968).

[2]B. C. Westmoreland, W. Szybalski, and H. Ris, Science, 163, 1343 (1969).

[3]B. Hudson, W. B. Upholt, J. Devinny, and J. Vinograd, Proc. Nat. Acad. Sci., 62, 813 (1969).

[4]D. A. Clayton, R. W. Davis, and J. Vinograd, J. Mol. Biol., in press.

[5]B. L. McConaughy, C. D. Laird, and B. J. McCarthy, Biochemistry, 8, 3289 (1969).

[6]H. R. Massie and B. H. Zimm, Biopolymers, 7, 475 (1969).

[7]J. G. Wetmur and N. Davidson, J. Mol. Biol., 31, 349 (1968).

[8]W. B. Studier, J. Mol. Biol., 71, 199 (1969).

[9]A. K. Kleinschmidt and R. K. Zahn, Z. Naturforsch., 14B, 770 (1959).

[10]A. K. Kleinschmidt, Methods in Enzymology, Vol. XIIB, p. 125 (1968); L. Grossman and K. Moldave, eds.; Academic Press, New York and London.

[11]H. Bujard, Proc. Nat. Acad. Sci., U.S., 62, 1167 (1969).

[12]J. G. Wetmur, N. Davidson, and J. V. Scaletti, Biochem. Biophys. Res. Commun., 25, 684 (1966).

Footnotes (continued)

[13]C. N. Gordon and A. K. Kleinschmidt, Biochim. Biophys. Acta, 155, 305 (1968).

[14]C. A. Thomas and L. A. MacHattie, Proc. Nat. Acad. Sci., 52, 1297 (1964).

[15]C. S. Lee, R. W. Davis, and N. Davidson, J. Mol. Biol., submitted.

[16]R. B. Inman, J. Mol. Biol., 25, 209 (1967).

[17]R. L. Sinsheimer, J. Mol. Biol., 1, 43 (1959).

[18]E. Burgi and A. D. Hershey, Biophys. J., 3, 309 (1963).

[19]L. A. MacHattie and C. A. Thomas, Jr., Science, 144, 1142 (1964).

[20]C. C. Richardson and B. Weiss, J. Gen. Physiol., 49, 81 (1966).

[21]F. W. Studier, J. Mol. Biol., 11, 373 (1965).

[22]J. S. Parkinson and R. W. Davis, manuscript in preparation.

[23]Z. Hradecna and W. Szybalski, Virology, 32, 633 (1967).

[24]M. Simon, R. W. Davis, and N. Davidson, manuscript in preparation.

[25]C. S. Lee and N. Davidson, Virology, in press.

[26]K. Wulff, J. Jamieson, and N. Davidson, manuscript in preparation.

Fig. 1. Electron micrographs of heteroduplexes

(a) and (b). Heteroduplexes mounted by the aqueous technique and stained with uranyl acetate.

(c) and (d). Heteroduplexes mounted by the formamide technique and shadowed with platinum-palladium.

(a) $\lambda_{b522}/\lambda_{b536}$ heteroduplex. The small deletion bush (arrow) near the top of the micrograph is the b536 marker bush. Therefore, the left end of the molecule is at the top of the micrograph and the other larger deletion bush is the b522 bush. The mapping problem is discussed in the text.

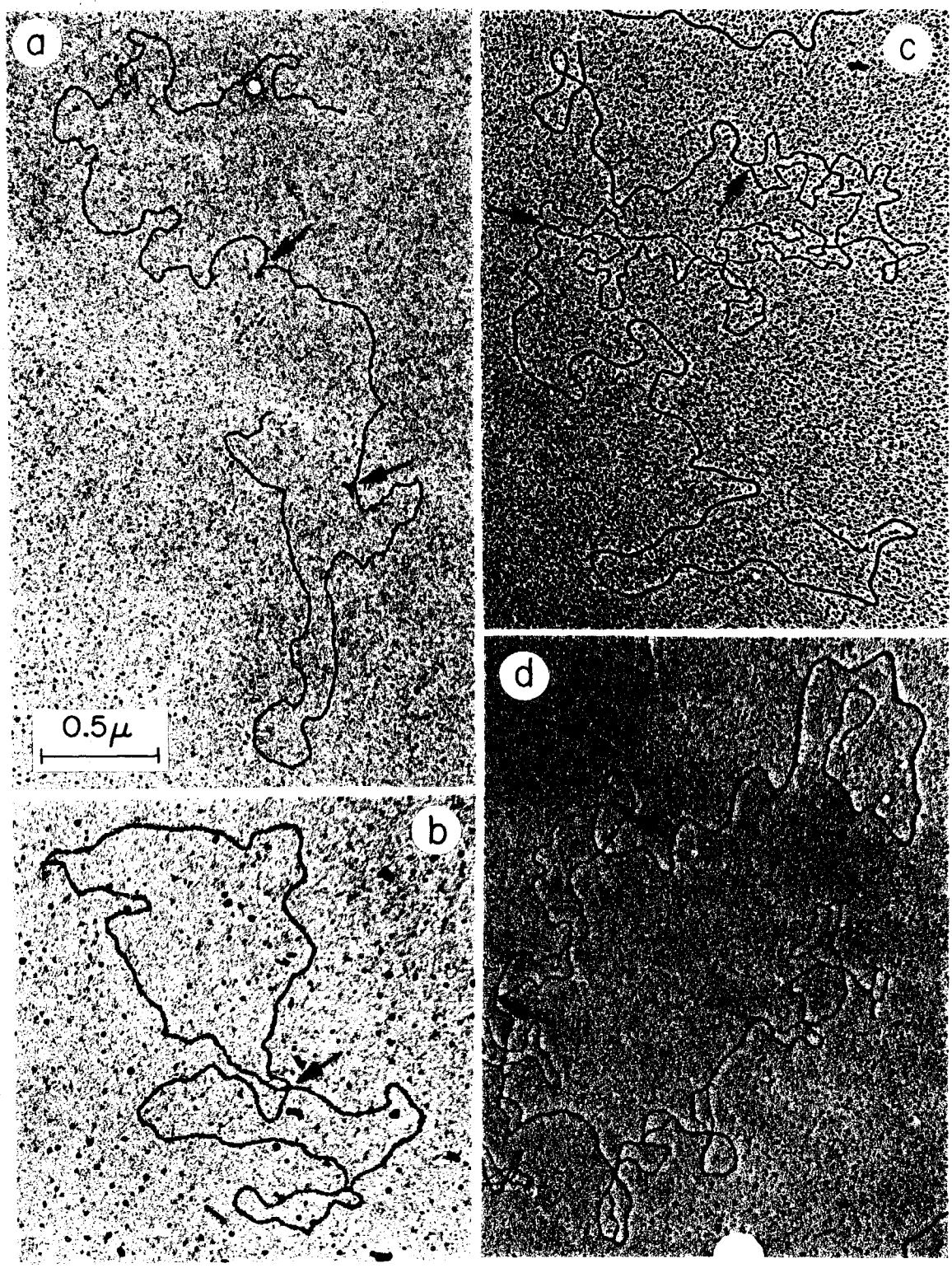
(b) Dimer/monomer mitochondrial DNA heteroduplex (fused dimer), (Clayton, Davis and Vinograd[4]). The mitochondrial DNA was prepared from a single untreated patient with granulocytic leukemia. The heteroduplex is composed of one dimer strand and two monomer strands. The molecule appears as a "figure 8" and is called a fused dimer. The fusion point between the two monomer length circles is marked by an arrow. The contour length of the two circles on either side of the fusion point are the same. No nonhomology regions between the dimer and monomer strands were observed. The originally closed circular DNA sample was nicked by the ethidium Br-light treatment and renatured by the microdialysis technique.

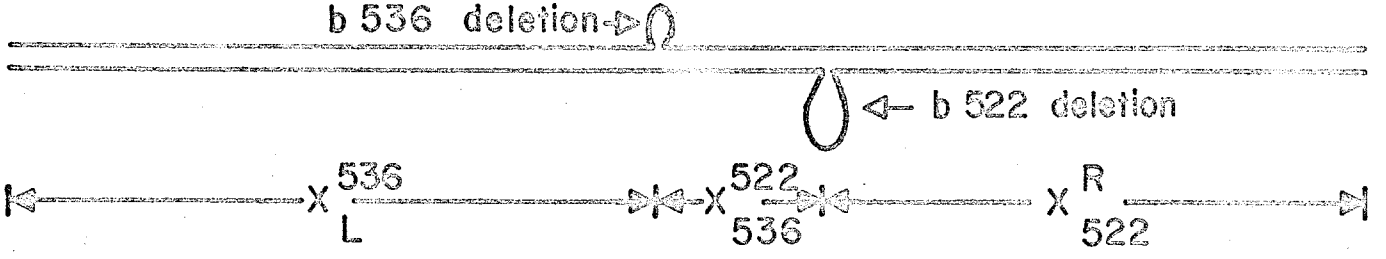
(c) λdg/$\lambda^{++}$ heteroduplex. In the λdg, a large segment of the λDNA (gene G to the attachment site at 0.574 λ lengths from the left end of the molecule) has been substituted by a segment of E. coli DNA (galactose genes). Therefore, the λdg/$\lambda^{++}$ heteroduplex shows a large region of non-homology. The two arrows mark the end points of the substitution.

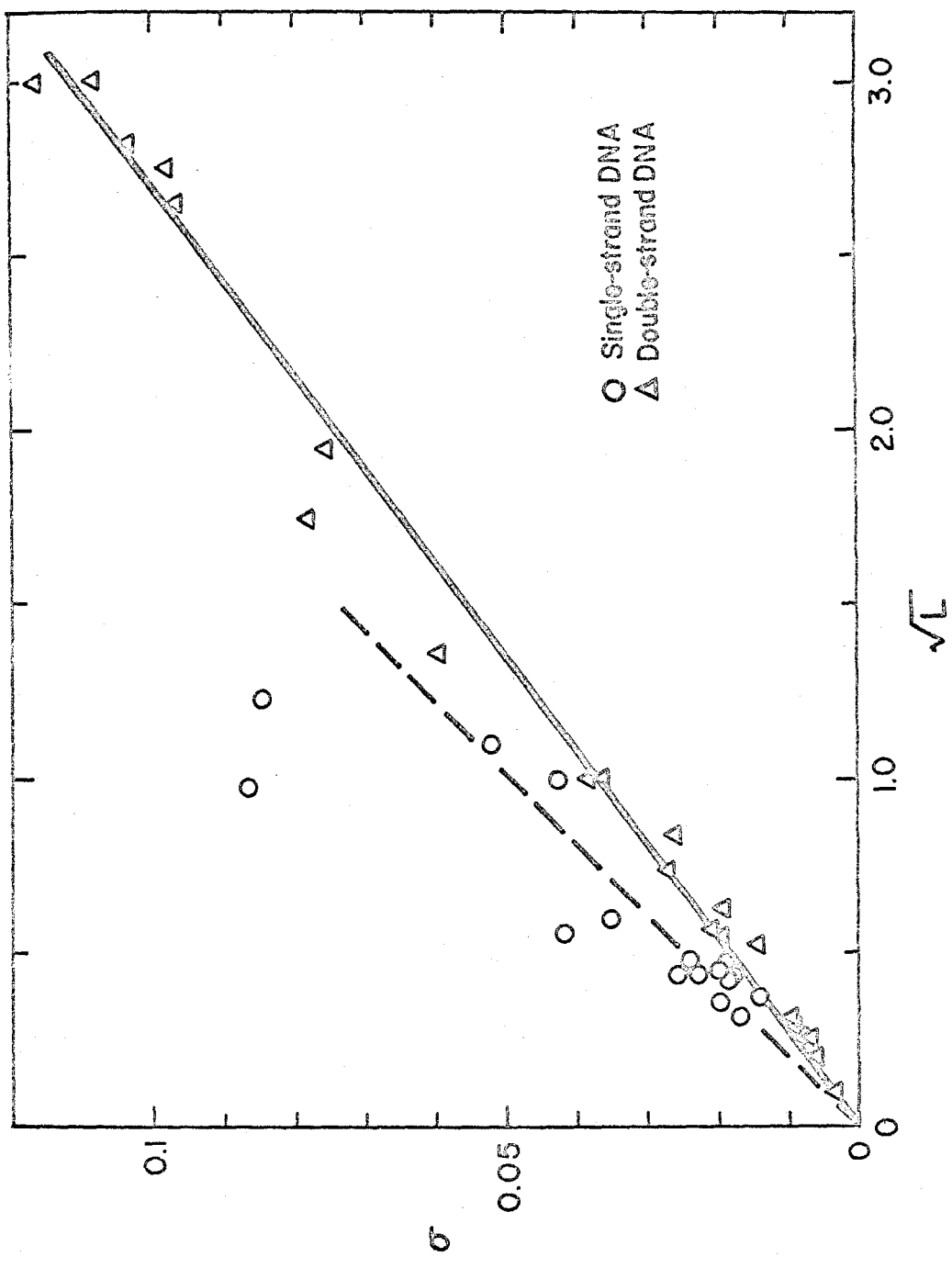
(d) 434/λ heteroduplex. This is a heteroduplex between two closely related bacteriophages. Note the large number (10) of nonhomology regions. The percent homology between these two phages as measured by electron microscopy is 37%. A striking feature of this heteroduplex is that the regions of homology and nonhomology are large and often the length of one or more genes. The arrows mark the two termini of the heteroduplex loop discussed in the text. The two single strands of this loop have lengths of 16% and 10% of λ.

Fig. 2. Schematic map of the $\lambda_{b522}/\lambda_{b536}$ heteroduplex. See Fig. 1a. The symbols are defined in the text.

Fig. 3. Plot of standard deviation of length measurements for samples of double- and single-strand DNA versus the square root of the mean length, $\langle L \rangle$. Most of the points represent samples of 25 or more molecules. The double-strand points include λ and some of its deletion mutants, E. coli minicircles, N phage DNA's, φX RF DNA, and various segments of λ heteroduplexes. Measurements were made on samples mounted by both the aqueous technique and the formamide technique.


a
0.5μ
b
c
d


b 536 deletion
b 522 deletion
X
536
L
X
522
536
X
R
522


Single-strand DNA
Double-strand DNA
σ
0.1
0.05
0
0
1.0
2.0
3.0
√L

## B. Separation of Lambda DNA Strands

The complementary strands of λ DNA can be preparatively separated by CsCl density gradient centrifugation of a mixture of denatured λ DNA and poly rG, poly rUG, or poly rIG (Hradecna & Szybalski, 1967). The procedure outlined by these authors was used to separate the strands of λ DNA. It was found that hybridization of the two, freshly prepared, strands resulted in the formation of substantial whole duplex DNA, while self-hybridization of each separated strand did not result in the formation of significant amounts of duplex DNA. However, on storage (two weeks in buoyant CsCl and 0.01 M EDTA at pH 8), the separated strands, prepared as described by Hradecna & Szybalski (1967), became fragmented and gave very little whole duplex DNA. It was also discovered that on alkaline velocity sedimentation through a sucrose gradient, the freshly prepared separated strands were very fragmented. Under conditions in which whole (unseparated) λ DNA strands sediment to the middle of the gradient, neither of the two freshly prepared separated strands sedimented away from the poly UG at the top of the gradient. It is believed that the heat denaturation (95° C for 2 min in $10^{-3}$ M $Na_3EDTA$) used by Hradecna & Szybalski (1967) results in the substantial depurination of the DNA. Chain scission is not immediate but does occur with time (on the order of

days) or when placed in alkali. Therefore, a procedure was developed which uses alkaline denaturation instead of heat denaturation. A detailed description is given here for the separation of the complementary strands of λ DNA by alkaline denaturation with poly rUG, by alkaline denaturation with poly rG, and by heat denaturation with poly rUG.

(i) Polyribonucleotides

The poly rG and poly rIG were from Biopolymers, Inc. The poly rUG was from Miles Laboratories, Inc. G:U=1:1.35.

(ii) DNA preparation

(a) Strand separation using alkaline denaturation with poly rUG.

1) 50 μl of λ phage with an $A_{260}$ = 20 in $10^{-2}$ M EDTA at pH 8 and anywhere from 0 to 1 M CsCl are placed in a Polyallomer centrifuge tube.
2) 4 μl of 5% sodium lauryl sarcosine (SDSar) in $10^{-2}$ M Tris at pH 7.5.
3) 6 μl of 1 N NaOH.
4) Ten min at room temp (23° C), then cool in ice.
5) Place on a vortex mixer (Scientific Products) at position 1.5.
6) Quickly add, at 0° C, 0.47 ml of 0.015 M Tris-HCl, 0.001 M Tris-base, 0.001 M EDTA, containing 50 μg of poly rUG.
7) Place at 5° C.

8) Quickly add 2.13 ml of saturated (at 23° C) CsCl to the DNA in the Polyallomer tube.

9) Fill the tube with mineral oil.

(b) Strand separation using alkaline denaturation with poly rG.

1) 50 μl of λ phage with an $A_{260}$ = 20 in $10^{-2}$ M EDTA at pH 8 and anywhere from 0 to 1 M CsCl are placed in a Polyallomer centrifuge tube.

2) 50 μl of poly rG, 1 mg/ml.

3) 4 μl of 5% SDSar.

4) 11 μl of 1N NaOH.

5) 10.0 min at room temp (23° C), then cooled in ice.

6) Place in a vortex mixer at position 1.5.

7) Quickly add at 0° C, 0.42 ml of 0.035 M Tris-HCl, 0.005 M Tris-base, $10^{-3}$ M EDTA.

8) Place at 5° C.

9) Quickly add 2.13 ml of sat CsCl to Polyallomer tube.

10) Fill the tube with mineral oil.

(c) Strand separation using heat denaturation and poly rUG.

1) 50 μl of phage with an $A_{260}$ = 20 in $10^{-3}$ M EDTA at pH 8 prepared by dialysis against $10^{-3}$ M EDTA.

2) 11 μl of poly rUG, 4.5 mg/ml ($A_{260}$ = 90)

3) 4 μl of 5% SDSar in $10^{-2}$ M Tris.

4) 0.185 ml $H_2O$ total volume = 0.25 ml.

5) Heat to 95° C for 2 min.

6) Cool quickly in ice water.

7) Add to 2.13 ml of sat CsCl in Polyallomer centrifuge tube.

8) Add 0.28 ml of $10^{-3}$ M EDTA.

9) Fill tube with mineral oil.

(iii) Centrifugation

Centrifugation was done in a Beckman Model L ultracentrifuge with a SW 39 or a SW 50 rotor at 30,000 rpm for 70 hr at 5° C.

(iv) Fractionation

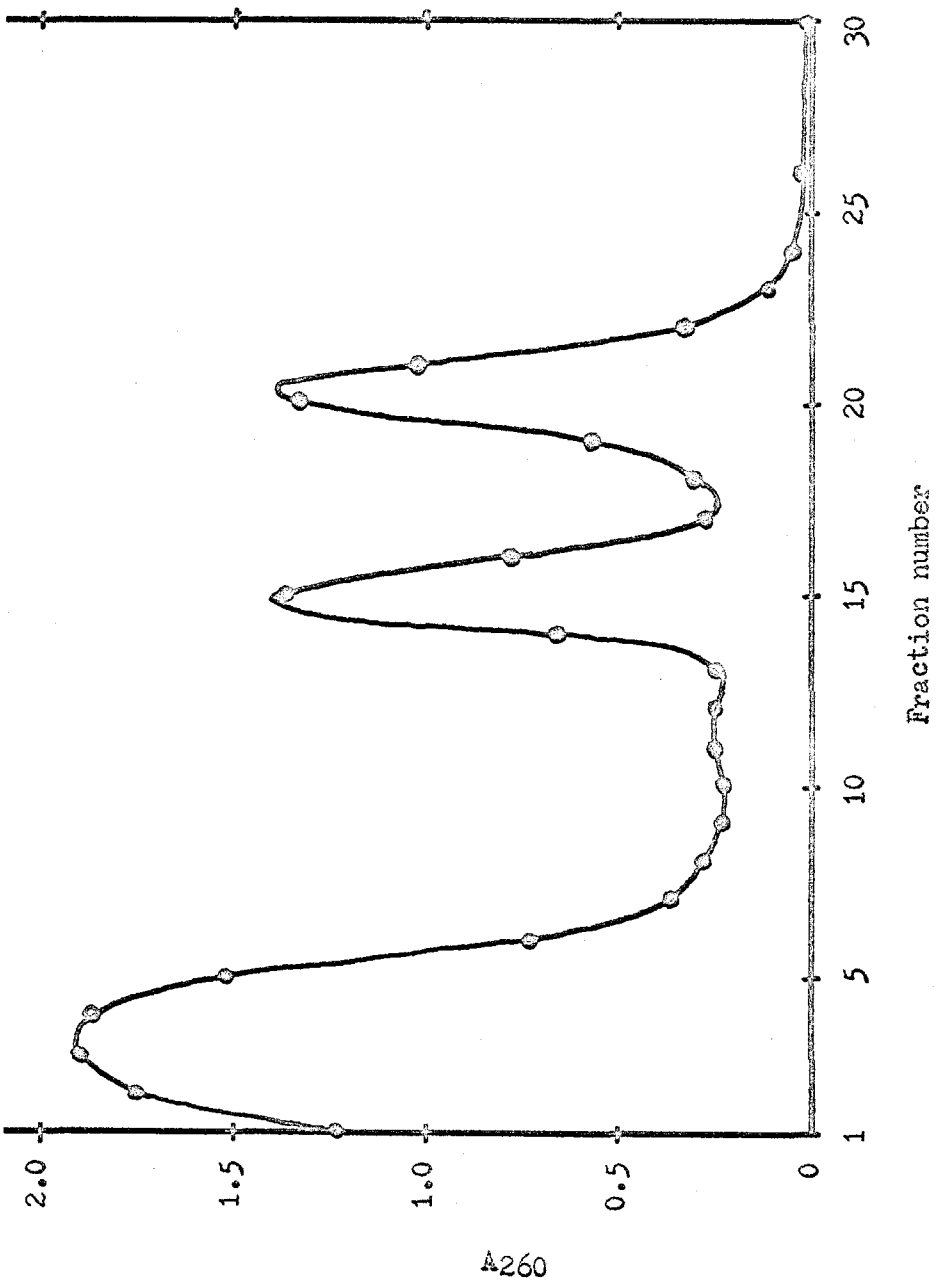
Ten drop fractions (90 to 100 $\mu$l) were taken with a Buchler fractionator and collected in a Teflon dripping block. Approximately 30 fractions were taken for each tube. The $A_{260}$ of each fraction was measured in a microcell, light path length = 1 cm, width = 2 mm, height = 4mm (tapped off at top and bottom) with a Cary 14 spectrophotometer. The results are shown in Fig. 1-4.

Reference

Hradecna, Z. and W. Szybalski (1967). Virology, 32, 633.

Fig. 1-4. Preparative separation of λ DNA strands.

The complementary strands of λ DNA were separated by the alkaline denaturation and poly rUG procedure as outlined in the text. The peak at fraction 3 is the band of excess poly rUG. If poly rG had been used, it would not have formed a band but would pellet on the bottom. The peak at fraction 15 is the r strand of λ DNA and the peak at fraction 20 is the l strand of λ DNA.


2.0
1.5
1.0
0.5
0
A260
1
5
10
15
20
25
30
Fraction number

# CHAPTER 2

## Electron-microscopic Visualization of Deletion Mutations

Ronald W. Davis and Norman Davidson

The following paper appeared in the Proceedings of the National Academy of Sciences, vol. 60, No. 1, pp. 243-250, May, 1968, and is reproduced here with the express written permission of the publisher.

Reprinted from the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES
Vol. 60, No. 1, pp. 243–250. May, 1968.

# *ELECTRON-MICROSCOPIC VISUALIZATION OF DELETION MUTATIONS**

BY RONALD W. DAVIS AND NORMAN DAVIDSON

GATES, CRELLIN, AND CHURCH LABORATORIES OF CHEMISTRY,
CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA

*Communicated February 12, 1968*

Deletion mutants of λ coliphage were discovered by Kellenberger, Zichichi, and Weigle.[1] The deletions can be mapped by genetic recombination experiments. It is not known if the recombination maps thus obtained give the true physical position of the deletions. The present paper describes a method for mapping deletion mutations with the electron microscope, thus obtaining the physical position of the deletion. The basic principle is as follows. A mixture of DNA molecules from wild-type virus and from the deletion mutant is subjected to strand dissociation followed by reannealing conditions. The resulting preparation contains some double-stranded renatured molecules of each type and some heteroduplexes. In the latter, one strand is wild-type DNA and the other is deletion mutant-type DNA. Each heteroduplex should thus contain a single-stranded loop in the wild-type DNA strand at the point where the deletion occurs. The contour lengths of double-stranded regions of a DNA molecule can be rather accurately measured in electron micrographs. Under the conditions used to prepare the electron microscope grids, the single-stranded loops are collapsed into "bushes" similar to those observed in $T_2$ phage DNA by MacHattie, Ritchie, Thomas, and Richardson.[2] The contour length of such a bush can only be roughly estimated at best, but the bush position with respect to the double-stranded regions can be accurately measured.

*Materials and Methods.*—The $c_I$ point mutant $\lambda c_{26}$ and the deletion mutants $\lambda b_2 b_5 c_1$, $\lambda b_5 c_2$ (Kellenberger, Zichichi, and Weigle[1]), and $\lambda b_{221} c_{26}$ (Huskey[3]) were obtained from Dr. Robert Huskey. Phage were grown and purified by standard methods, including banding in CsCl by density gradient centrifugation.[4] The phage solution was finally dialyzed against 0.01 *M* $MgSO_4$, 0.01 *M* tris buffer (pH 7).

A typical heteroduplex preparation was made as follows. Ten μl of 0.1 *M* ethylenediaminetetraacetate (EDTA) (pH 8.2), 15 μl of 5 *M* NaCl, 46.6 μl of $H_2O$, 5.1 μl of $\lambda c_{26}$ phage solution ($A_{260}$ = 10.0), 3.4 μl of $\lambda b_5 c_2$ solution ($A_{260}$ = 14.2), and 10 μl of 1 *N* NaOH were added, in the order given, to a small test tube (6 × 50 mm). The mixture (pH 13) was incubated at room temperature for 10 min and then chilled in an ice bath. The alkaline solution causes lysis of the phage and dissociation of the DNA into single strands. The solution was neutralized by addition of 10 μl of 2 *M* $NaH_2PO_4$ (final volume, 100 μl). The DNA was partially renatured by heating the solution to 70°C for 30 sec and then quickly cooled (these conditions were selected to give over 50% renaturation, minimal single-strand scissions, and minimal higher-order aggregation[5]). All other preparations were done similarly except that the volume of phage suspension added was varied to give equal numbers of the two phage types with a combined $A_{260}$ = 1.0 (*ca.* $10^{11}$ total phage). In control experiments only one type of phage DNA was present.

Grids for electron microscopy were prepared by the basic protein film technique[6] and stained with uranyl salts.[7] A solution containing 0.5 μg/ml of DNA and 0.1 mg/ml of cytochrome *c*, in 0.5 *M* ammonium acetate, 0.001 *M* EDTA (pH 7.9), was spread onto 0.15 *M* ammonium acetate (pH 6.5). The cytochrome *c*–DNA-mixed film was picked up on grids freshly covered (less than 2-days old) with films prepared from 3% w/v

Parlodion in amyl acetate. Best staining results were obtained when the droplet of solution that was picked up with the cytochrome *c*–DNA film gave a flat meniscus on the Parlodion grid. The grids were stained by dipping them for 30 sec in a 90% ethanol solution of stain and then rinsed for 10 sec in isopentane. The best stain solution was a 1000-fold dilution into 90% ethanol of a stock solution of either 0.05 *M* uranyl chloride (made from uranyl oxide) in 0.05 *N* HCl or 0.05 *M* uranyl acetate in 0.05 *N* HCl. The stock solutions were stored in the dark. A fresh, diluted stain solution was prepared for each set of grids and was used within 1 hr after dilution. Some grids were also shadowed with platinum-palladium. DNA molecules were photographed with a Phillips EM 200 electron microscope using a 50-$\mu$ objective aperture and 60-kv accelerating voltage. The resulting negatives (approx. 3600X) were enlarged ten times and traced on a Nikon shadowgraph. Lengths were measured on the tracings with a map measurer. Fifteen to 40 molecules were measured for each heteroduplex and each control. Bush sizes were estimated by enlarging (50X) the micrographs on the shadowgraph and tracing the periphery of the bush.

The size of a deletion was determined with the electron microscope by a direct comparison of the length of wild-type and deletion mutant DNA's on the same grid. This procedure virtually eliminates magnification errors and differential environmental effects on the length of DNA. Phenol-extracted native DNA was used for each determination One comparison was also made using heat-shocked phage (300-fold dilution of phage stock into 0.001 *M* EDTA, 10 min at 37°C). Since the $b_5$ deletion is small, the histograms of $\lambda b_5 c_2$ and $\lambda c_{26}$ might overlap. Therefore, in order to identify the type of DNA in the electron microscope, circular (along with some linear) $\lambda c_{26}$ DNA was compared to linear $\lambda b_5 c_2$ DNA, and circular (along with some linear) $\lambda b_5 c_2$ DNA was compared to linear $\lambda c_{26}$ DNA. Grids were stained with uranyl chloride and shadowed with platinum-palladium (to prevent the Parlodion film from stretching).

*Results and Discussion.*—Ideally, when a mixture of two DNA's, one of them differing from the other by a genetic deletion, is denatured by alkali, neutralized, and heat-annealed, a mixture containing each of the two original double-stranded DNA's and a heteroduplex double-stranded DNA is obtained. This heteroduplex has, at the point of the deletion, a single-stranded bush which can be readily identified by electron microscopy.

Before presenting the results, certain difficulties in interpretation due to the formation of molecules with bushes that are not due to a deletion should be mentioned. These cases are probably due to imperfect annealing and/or to the occurrence of single-strand breaks. Some completely denatured single-strand molecules are seen but are readily identified (Fig. 2*a*, arrow). In a typical preparation approximately 33 per cent of the renatured molecules showed bushes on one or both ends. We believe that this end-bush formation results from the renaturation of two strands, of which one or both is incomplete due to a prior single-strand break. The double-stranded portion of such a molecule was always found to be shorter than a whole double-stranded molecule. For a duplex containing one unbroken and one broken strand, the estimated length of the bush (or bushes) should increase linearly with decreasing length of the double-stranded region. The correlation between measured periphery lengths of end bushes and lengths of the double-stranded regions is displayed in Figure 1. The data give the predicted linear relation and a conversion factor of 1/3.2 between bush periphery length and double-stranded length. The single-strand breaks were mapped (by the length of the double-stranded region) and appear to be randomly located.

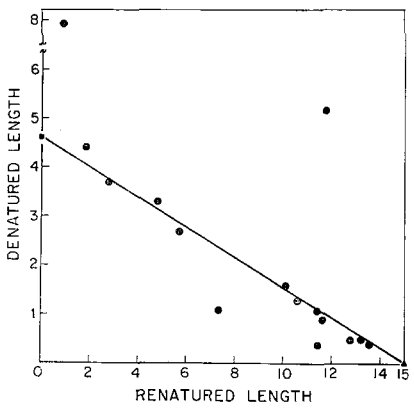



FIG. 1.—The lengths of the double-stranded portions of renatured molecules with end bushes are plotted against the lengths of the peripheries of the end bushes. The straight line is drawn between whole single-stranded DNA (■) (mean of eight measurements) and whole double-stranded DNA (▲) (mean of 31 measurements). It is seen that the ratio of the double-stranded contour length of a whole molecule to the single-stranded periphery length of a whole strand is 3.2. The two points far above the line are probably the result of aggregation of a whole single strand with an end bush. The two points below the line are probably the result of annealing two fragmented single strands.

Considering now only the renatured molecules without end bushes, in all heteroduplex preparations approximately half the molecules showed internal bushes, while in control preparations less than 5 per cent of the molecules showed internal bushes. The internal bushes on renatured molecules in the control preparations (that is, from a single kind of DNA) are probably mainly due to nonspecific aggregation of a single-stranded segment with a renatured two-stranded molecule or to hybridization of, for example, a broken $W$ strand with a whole $C$ strand, followed by renaturation with an additional whole $W$ strand. The internal bushes in the control preparations appear to be randomly located. As noted below, the internal bushes that are identified as a deletion always occur at a unique location.

In heteroduplex preparations, pictures were taken of all isolated molecules which showed internal bushes and which were sufficiently straightened out so that intramolecular overlaps caused no ambiguity as to the position of the bush. Typical molecules are shown in Figure 2.

The measured positions of the bushes in the heteroduplexes can be used to calculate the positions of the deletions in the wild-type DNA if the size of the deleted region is independently known. This calculation proceeds as follows. Let $L_0$ be the physical length of the wild-type molecule, with $X_1^i$ and $X_2^i$ the physical positions (measured from the left end) of the beginning and end of the $i$th region, which is missing in the deletion mutant, $i$. The physical length of the deletion is $D^i = X_2^i - X_1^i$. The fractional or map positions are $x_1^i = X_1^i/L_0$ and $x_2^i = X_2^i/L_0$, and the fractional size of the deletion is $d^i = D^i/L_0 = x_2^i - x_1^i$. The deletion-mutant molecule has a length $L_0 - D^i$; the total length of the double-stranded regions of a heteroduplex between a wild-type and an $i$-type molecule is also $L_0 - D^i$. Therefore, the physical distance from the left end of the heteroduplex to the bush due to region $i$ is $X_1^i$; the fractional length is $X_1^i/(L_0 - D^i)$ or $x_1^i/(1 - d^i)$. In a heteroduplex between a molecule with deletions of the two nonoverlapping regions $i$ and $j$ and a molecule from which only $i$ is missing, the bush due to the $i$th region occurs at $x_1^i/(1 - d^i - d^j)$ or at $(x_1^i - d^j)/(1 - d^i - d^j)$, depending on whether the $i$th region is to the left or the right of the $j$th region.

In a study of the molecules from the $b_5$ heteroduplex preparation ($\lambda b_5 c_2 \times \lambda c_{26}$) (Fig. 2$c$), the internal bushes were found to be uniquely located as shown by

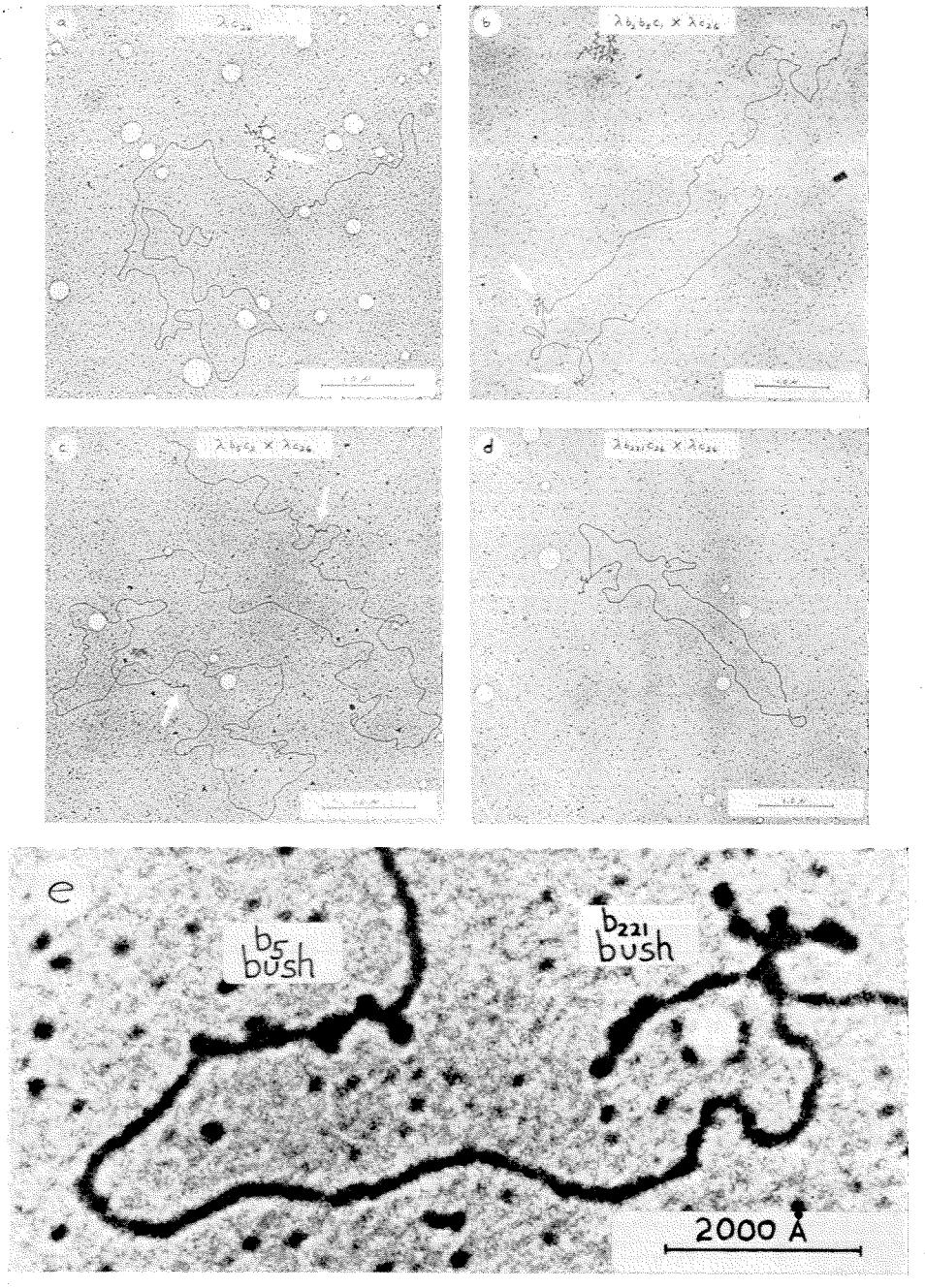



FIG. 2.—Electron micrographs of renatured λ DNA, stained with uranyl salts.
(*a*) $\lambda c_{26}$ DNA self-annealed. Arrow shows a single strand of DNA.
(*b*) $\lambda b_2 b_5 c_1 \times \lambda c_{26}$ heteroduplex showing two bushes (*arrows*).
(*c*) Two $\lambda b_5 c_2 \times \lambda c_{26}$ heteroduplexes, a circular and a linear molecule, each showing one small bush (*arrows*).
(*d*) $\lambda b_{221} c_{26} \times \lambda c_{26}$ heteroduplex showing one large bush and showing an apparently pulled-apart circle.
(*e*) Enlarged micrograph showing the $b_5$ and $b_{221}$ bushes.

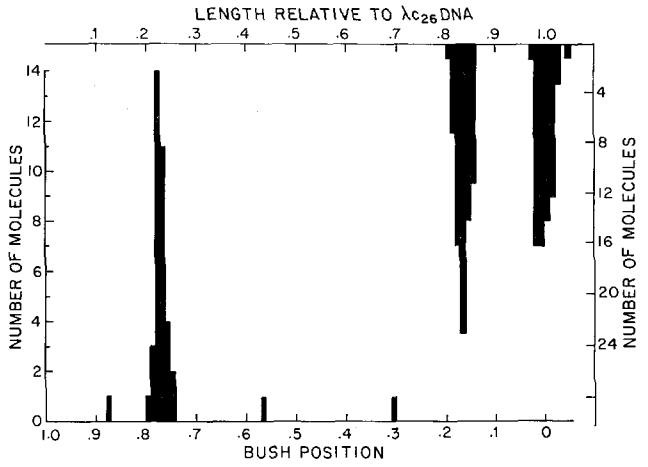



FIG. 3.—Histogram of bush positions for molecules with internal bushes in a $\lambda c_{26} \times \lambda b_5 c_2$ heteroduplex preparation. The horizontal coordinate is the distance of a bush from an end, measured as a fraction of the renatured molecular length, and is therefore the quantity $x_1^{b5}/(1 - d^{b5})$.

Histogram of the relative lengths for a mixture of $\lambda c_{26}$ and $\lambda b_2 b_5 c_1$ DNA molecules on the same grid. The mean of the $\lambda c_{26}$ DNA lengths is taken as 1.00. The standard deviation of the length measurements is 1.2%.

the histogram in Figure 3. Ignoring the three randomly located bushes, we find the mean position of the other 35 to be $x_1^{b5}/(1 - d^{b5}) = 0.770$ with a standard deviation of $\pm$ 0.009. Since the $b_5$ deletion alters the product of the $c_I$ gene, one might assume that this deletion involves or is near the $c_I$ gene.[8] The $c_I$ gene genetically maps in the right quarter of the $\lambda$ DNA molecule, and this is in good agreement with the bush position if the left end is taken as 0 and the right end as 1 unit of length.

A better deletion to study is the $b_2$ deletion, since there is good genetic evidence that this deletion is in the center of the $\lambda$ DNA molecule.[9] However, it is not known if this deletion is slightly left or right of center; therefore, the double deletion $b_2$ and $b_5$ heteroduplex was prepared ($\lambda b_2 b_5 c_1 \times \lambda c_{26}$). As expected, this heteroduplex shows two internal bushes (Fig. 2*b*). These internal bushes are uniquely located at 0.736- and 0.552-molecular lengths (one molecular length now being that of the double deletion $b_2 b_5$ heteroduplex DNA molecule). If these two bushes are the $b_5$ and $b_2$ regions, respectively, then $(x_1^{b5} - d^{b2})/(1 - d^{b5} - d^{b2}) = 0.736$ and $x_1^{b2}/(1 - d^{b5} - d^{b2}) = 0.552$.

To show that the bush at 0.552 molecular length is the $b_2$ bush, the $\lambda b_2 b_5 c_1 \times \lambda b_5 c_2$ heteroduplex was prepared. This heteroduplex should show only the $b_2$ bush, since both DNA strands have the $b_5$ deletion. A single bush is observed which maps at $x_1^{b2}/(1 - d^{b5} - d^{b2}) = 0.544$.

If we take $d^{b2} = 0.112$ and $d^{b5} = 0.053$ (see below), the observation that $x_1^{b5}/(1 - d^{b5}) = 0.770$ implies that $x_1^{b5} = 0.729$. This predicts that $(x_1^{b5} - d^{b2})/(1 - d^{b2} - d^{b5}) = 0.738$, in satisfactory agreement with the observed value, 0.736. The position of the $b_2$ deletion is calculated as $x_1^{b2} = 0.458$. Then $x_2^{b2} = 0.570$; and the $b_2$ region is seen to be very close to the center of the molecule.

To compare further genetic and DNA distance mapping, a new, large deletion mutation, $b_{221}$, isolated by Dr. Robert Huskey, was studied. His genetic studies[3] show that this deletion overlaps the $b_2$ deletion and extends farther on both sides of it. The $b_{221}$ heteroduplex ($\lambda b_{221} c_{26} \times \lambda c_{26}$) gives a single, large internal

TABLE 1. *Length and bush position for renatured molecules, and deletion region coordinates.*

| Renatured DNA | Length in $\mu$ | Bush position in fractions of the molecular length | Deletions Regions in Fractions of the Wild-Type DNA Length Measured from Left End | |
|---|---|---|---|---|
| $\lambda c_{26}$ | 15.0 ± 0.4 | | | |
| $\lambda b_5 c_2$ | 13.0 | | | |
| $\lambda b_5 c_2/\lambda c_{25}$ | 13.3 | 0.770 | $x_1^{b5} = 0.729$ | $x_2^{b5} = 0.782$ |
| $\lambda b_2 b_5 c_1$ | 11.7 | | | |
| $\lambda b_2 b_5 c_1/\lambda c_{26}$ | 11.6 | 0.552 | $x_1^{b2} = 0.461$ | $x_2^{b2} = 0.573$ |
| | | 0.736 | $x_1^{b5} = 0.727$ | $x_2^{b5} = 0.780$ |
| $\lambda b_2 b_5 c_1/\lambda b_5 c_2$ | 11.7 | 0.544 | $x_1^{b2} = 0.454$ | $x_2^{b2} = 0.566$ |
| $\lambda b_{221} c_{26}$ | 11.2 | | | |
| $\lambda b_{221} c_{26}/\lambda c_{26}$ | 11.2 | 0.523 | $x_1^{b221} = 0.406$ | $x_2^{b221} = 0.629$ |
| $\lambda b_{221} c_{26}/\lambda b_5 c_2$ | 9.8 | 0.568 | $x_1^{b221} = 0.411$ | $x_2^{b221} = 0.634$ |
| | | 0.703 | $x_1^{b5} = 0.732$ | $x_2^{b5} = 0.785$ |

bush at 0.523-molecular length from an end (Fig. 2*d*). In order to locate this deletion, left or right of center, the double-deletion heteroduplex ($\lambda b_{221} c_{26} \times \lambda b_5 c_2$) was studied. This heteroduplex gives two internal bushes (Fig. 2*e*), one mapping at 0.703 and the other at 0.568-molecular lengths. Calculation then shows that the $b_{221}$ deletion is centered in the right half of the $\lambda$ DNA molecule, with $x_1 = 0.408$ and $x_2 = 0.631$. Thus, this deletion overlaps the $b_2$ deletion, in agreement with genetic mapping.

One further control is that the deletion mutant strand of DNA in a heteroduplex should define the length of the heteroduplex DNA molecule so that the heteroduplex molecule should have the same length as the self-annealed deletion mutant DNA molecule. That this is the case is demonstrated in Table 1. It should be pointed out that good estimates of the per cent deletion cannot be made from these length measurements, since the conditions which cause fluctuations in length were not rigidly controlled.[10] Good estimates of the per cent deletion were obtained, however, by mounting both deletion mutant and wild-type DNA's (phenol-extracted native DNA) on the same grid. As shown in Figure 3, sharp symmetrical histograms were obtained with a standard deviation of about 1 per cent. The per cent deletion was also determined, using the buoyant densities of the whole phage.[11] All these data are presented in Table 2. It is seen that the per cent deletion calculated from buoyant density agrees

TABLE 2. *Per cent deletion by electron microscopy and by buoyant density.*

| DNA | Form | Per cent deletion elec. micro.* | Method of preparation | No. measured | Buoyant density of phage† | Per cent deletion by buoyant density‡ |
|---|---|---|---|---|---|---|
| $\lambda b_5 c_2$ | Linear | 5.3 | Phenol-extracted | 54 | 1.502 | 5.5 |
| $\lambda b_5 c_2$ | Circular | 5.2 | Phenol-extracted | 59 | — | — |
| $\lambda b_2 b_5 c_1$ | Linear | 16.5 | Phenol-extracted | 99 | 1.487 | 17.9 |
| $\lambda b_2 b_5 c_1$ | Linear | 16.5 | Heat-shocked | 36 | — | — |
| $\lambda b_{221} c_{26}$ | Linear | 22.3 | Phenol-extracted | 82 | 1.480 | 23.1 |

* Per cent shorter than $\lambda c_{25}$ linear (or circular) DNA.

† Buoyant density in CsCl of $\rho^{26} = 1.493$, taking the buoyant density of the $\lambda c_{25}$ marker as 1.508.

‡ Per cent deletion $= -100\Delta\rho/(0.106 - 0.54\ \Delta\rho)$. This equation follows from equation (1) of Weigle *et al.*,[11] by taking $\rho_{DNA} = 1.704$ and $\rho_{protein} = 1.28$. The quantity $\Delta\rho$ is the buoyant density of the mutant phage minus that of $\lambda c_{26}$ phage and is negative, corresponding to a positive value of the per cent deletion.

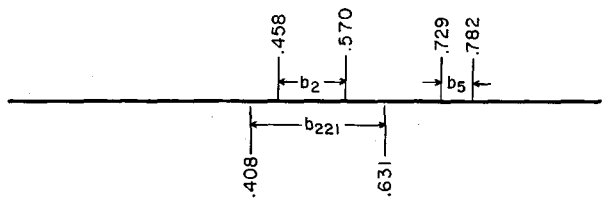



FIG. 4.—Summary of results for positions of the $b_{221}$, $b_2$, and $b_5$ regions.

rather well with that measured by electron microscopy. We estimate the standard deviation in our determination of deletion size as ±0.02-molecular lengths. In general, these measured deletion sizes are slightly smaller than those reported in a number of other investigations by several physical methods, but the discrepancies appear to be within the estimated respective errors. The values for $x_1$ and $x_2$ for the several deletions studied, as quoted above, were calculated using the values for the fractional deletion size ($d$) measured by electron microscopy (Table 2). These results are summarized in Table 1 and Figure 4.

The results of the present physical mapping study can be compared with results obtained in several other studies. Skalka, Burgi, and Hershey[12] have deduced the positions of segments of differing base composition along the λ DNA molecule by a physical method. They conclude that there is a segment of low GC mole fraction (0.37) extending from 0.44 to 0.54 of the λ DNA molecule and largely missing in λ$b_2$ mutants. The position of the $b_2$ region, as determined by us, is from 0.455 to 0.567. These results support the inference that the $b_2$ region and the segment of low GC content are virtually identical. The position of the *N* gene has been determined by a method based on the infectivity of fragments of different sedimentation coefficients to be 0.27 molecular lengths from the right end of the λ DNA molecule.[13] The *N* gene is close to the left end of the $b_5$ region, which we map at 0.273 lengths from the right end of the molecule.

Genetic recombination frequencies between the boundaries of the $b_2$ region and various markers to the left and the right have been studied by several investigators. Similarly, mutations in the $c_I$ gene, which is in the $b_5$ region, have been genetically mapped.[3, 8, 9, 14–16] All of these results are in semiquantitative agreement, at least, with the physical data presented here. A quantitative critical collation of the genetic recombination data is beyond the scope of the present contribution.

In the course of these experiments several interesting incidental observations have been made. All preparations contained some hydrogen-bonded circular molecules. In several cases, ends which were close and which appear to have been pulled apart during or after incorporation in the cytochrome film were observed in both heteroduplex and self-annealed molecules (Fig. 2*d*). The "break" always occurred at the cohesive site as determined by bush position. The two ends are not connected by single-stranded DNA, for single-stranded DNA can be seen by the grid preparation technique used.

Another observation concerns the appearance of the $b_5$ bush. The bush formed by a single continuous deletion should originate from a single point. This is true for all the $b_2$ and $b_{221}$ bushes that were observed. However, the $b_5$ bush

does not appear at one point; rather, it is usually spread over a region (the longest regions were 0.25-$\mu$ long) with up to five small "bushlets." These phenomena are illustrated in Figure 2*e*. Out of 122 well-stained $b_5$ bushes studied at high magnification, only eight appeared as a single bush, while all others appeared as multiple bushlets. Attempts were made to map the multiple bushlets, but no unique positions could be found. These characteristics of the $b_5$ bush indicate that the $b_5$ deletion is not a simple continuous deletion. In view of the altered immunity caused by the $b_5$ mutation,[8] it would appear that the $b_5$ deletion is really a DNA substitution, in which the substituted piece contains 5.3 per cent less DNA and the substitution shows partial homology with wild-type $\lambda$ DNA.

It is anticipated that the method described here will be useful for a number of problems involving nucleic acid homology.

*Summary.*—Heteroduplex DNA molecules in which one strand is from the wild-type viral DNA and the other strand from the DNA of a deletion mutant strain can be prepared by renaturation. These molecules consist mainly of double-stranded regions and are readily visualizable in the electron microscope by the basic protein film method. They show a bush (collapsed single-stranded loop) at the point where the deletion occurs. The bush positions can be accurately measured. The size of the deletion can be measured from the difference in contour length of the native wild-type DNA molecules and native deletion mutant DNA molecules. These two measurements permit a physical mapping of the position of the deletion regions. The $b_2$, $b_5$, and $b_{221}$ deletions in $\lambda$ DNA have been so mapped and the final results are presented in Figure 4.

After completing this manuscript we have learned that the technique described herein has been independently conceived and developed by Westmoreland, Szybalski, and Ris, who have obtained results which are generally similar to ours.

We are grateful to Dr. Robert Huskey for his help and phage stocks, to Dr. James G. Wetmur for advice on annealing conditions, and to Dr. Jean Weigle for his patient guidance and counsel in many matters. This work has been supported by research grant GM 10991 and training grant GM 1262 from the U.S. Public Health Service.

* Contribution no. 3649.

[1] Kellenberger, G., M. L. Zichichi, and J. Weigle, *J. Mol. Biol.*, **3**, 399 (1961).

[2] MacHattie, L. A., D. A. Ritchie, C. A. Thomas, Jr., and C. C. Richardson, *J. Mol. Biol.*, **23**, 355 (1967).

[3] Huskey, R. J., Ph.D. thesis, California Institute of Technology (1967).

[4] Wang, J. C., and N. Davidson, *J. Mol. Biol.*, **15**, 111 (1966).

[5] Wetmur, J. G., and N. Davidson, *J. Mol. Biol.*, **31**, 349 (1968).

[6] Kleinschmidt, A. K., and R. K. Zahn, *Z. Naturforsch.*, **14b**, 770 (1959).

[7] Wetmur, J. G., N. Davidson, and J. V. Scaletti, *Biochem. Biophys. Res. Commun.*, **25**, 684 (1966).

[8] Weigle, J., personal communication.

[9] Jordan, E., *J. Mol. Biol.*, **10**, 341 (1964).

[10] Inman, R. B., *J. Mol. Biol.*, **25**, 209 (1967).

[11] Weigle, J., M. Meselson, and K. Paigen, *J. Mol. Biol.*, **1**, 379 (1959).

[12] Skalka, A., E. Burgi, A. D. Hershey, *J. Mol. Biol.*, in press.

[13] Hogness, D. S., W. Doerfler, J. B. Egan, and L. W. Black, in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 31 (1966), p. 129.

[14] Franklin, N. C., *Genetics*, **57**, 301 (1967).

[15] Amati, P., and M. Meselson, *Genetics*, **51**, 369 (1965).

[16] Huskey, R. J., and J. S. Parkinson, personal communication.

# CHAPTER 3

## A Physical Study of the Integration of λ DNA into E. coli DNA

Ronald W. Davis and John S. Parkinson

The following paper has been prepared for submission to the Journal of Molecular Biology

## ABSTRACT

Deletion mutants and substitution mutants of phage λ have been used to examine the physical structure of the $\underline{att}^{\varphi}$ site in the λ chromosome which is essential for prophage integration. Integration-defective λ mutants were analyzed by constructing heteroduplex DNA molecules containing one wild type and one mutant strand and examining these heteroduplexes by electron microscopy. The results indicate that $\underline{att}^{\varphi}$ is less than 2500 base pairs in length and may be as small as 20 - 50 base pairs. Integration crossovers between $\underline{att}^{\varphi}$ and its bacterial analog , $\underline{att}^{B}$, occur within a region which is less than 20 base pairs in length. Moreover, there is no detectable base sequence homology between $\underline{att}^{\varphi}$ and $\underline{att}^{B}$. These results suggest that λ integration is a truly unique system of recombination.

# 1. INTRODUCTION

Bacteriophage lambda has evolved a highly specific recombination system for inserting its own DNA molecule into the continuity of the bacterial chromosome during lysogenization. The lambda integration machinery has at least two components, a functional one and a structural one. The functional part of this system is the Int gene (Zissler, 1967) whose product is essential for integration. It is not yet clear what role the Int product plays during integration, however, the Int product is assumed to be an integration enzyme or "integrase." The structural component of the integration system, called $\underline{att}^{\varphi}$, lies adjacent to the Int gene in the middle of the lambda genome. This phage attachment region is the site of action of the integrase and therefore must contain specific information necessary for recognition by the integration system.

The phage attachment site has at least three distinct components (Parkinson, 1970): a cross-over point (XOP) which is the site at which integrative recombination takes place, and a recognition element on each side of XOP. These elements are designated REL (recognition element left) and RER (recognition element right). While they influence the efficiency of integrative recombination, these elements are probably not uniquely essential for such recombination. In this paper we describe studies of the physical structure of $\underline{att}^{\varphi}$. In particular, we examined the following aspects of the physical organization of the attachment site:

(1) The size and location of XOP.

(2) The size and location of REL and RER.

(3) The extent of base sequence homology between $\underline{att}^{\varphi}$ and its bacterial analog $\underline{att}^{B}$.

In order to study att$^{\varphi}$ which seems to have a purely structural role, it is necessary to employ mutants such as deletions or substitutions which contain major structural changes of the att region. The isolation and genetic properties of such att-defective mutants are described in the first two papers of this series (Parkinson & Huskey, 1970; Parkinson, 1970). For the work described in this paper we have employed the technique of electron microscopy of heteroduplex DNA molecules (Davis & Davidson, 1968; Westmoreland, Szybalski & Ris, 1969) to examine the physical properties of att-defective mutants. By comparing the genetic properties of these mutants with their physical size and location in the lambda chromosome, we have obtained a detailed picture of att$^{\varphi}$ and of its role in the integration process.

## 2. MATERIALS AND METHODS

### (a) Phage strains

The phage mutants used in this work are listed in Table 1. All of these phages are derived from $\lambda^{++}$ ($\lambda$PaPa) of Kaiser (1957). The deletion mutants in this table are described by Parkinson & Huskey (1970) and by Parkinson (1970).

### (b) Bacterial strains

The bacterial strains employed in this study are described in Table 2.

TABLE 1

Phage Strains

a) Miscellaneous strains

| Strain | Relevant Properties | Source and/or reference |
|---|---|---|
| $\lambda^{++}$ | wild type | Kaiser (1957) |
| λint29 | int⁻ sus$_{21}$ | E. Signer |
| λb5 | imm | J. Weigle; Kellenberger, et al. (1961b) |

b) Deletion mutants (Parkinson and Huskey, 1970; Parkinson, 1970)

| Class of mutant | Relevant Properties[1] | Mutant isolates employed |
|---|---|---|
| 1 | int$^+$ att$^{\varphi}$ | b501, b502, b504, b506, b509, b510, b515, b519, b520, b536 |
| 2A | int$^+$ REL$^{\Delta}$ | b2[2], b511, b516, b527, b540 |
| 2B | int$^+$ REL$^B$ | b130, b189 |
| 3A | Int$^{\Delta}$ RER$^{\Delta}$ | b508, b512, b514, b517, b522 |
| 3B | Int$^{\Delta}$ att$^{\Delta}$ | b221, b531, b538 |

c) Transducing phages

| λdg isolate | Relevant Properties | Source and/or reference |
|---|---|---|
| P71 | (A-J)$^{\Delta}$ att$^L$ | isolated in this laboratory |
| P73 | (C-J)$^{\Delta}$ att$^L$ | " |
| P74 | (G-J)$^{\Delta}$ att$^L$ | " |
| P72 | (H-J)$^{\Delta}$ att$^L$ | " |
| **λbio isolate** | | |
| bio16A | Int$^{\Delta}$ att$^R$ | K. Manly; Manly, et al. (1969) |
| bio7-20 | (Int-Xis)$^{\Delta}$ att$^R$ | K. Manly; Kayajanian (1968) |
| bio69 | (Int-Exo)$^{\Delta}$ att$^R$ | K. Manly; Manly, et al. (1969) |
| bio11 | (Int-β)$^{\Delta}$ att$^R$ | " |
| bio10 | (Int-CIII)$^{\Delta}$ att$^R$ | " |

TABLE 1 (Cont'd)

(1) Abbreviations:

| | | |
|---|---|---|
| $\underline{att}^{\varphi}$ | = | wild type phage attachment site |
| $\underline{att}^{L}$ | = | left prophage attachment site |
| $\underline{att}^{R}$ | = | right prophage attachment site |
| $REL^{\Delta}$ | = | deletion of left recognition element in $\underline{att}^{\varphi}$ |
| $RER^{\Delta}$ | = | deletion of right recognition element in $\underline{att}^{\varphi}$ |
| $REL^{B}$ | = | substitution of bacterial DNA for left recognition element in $\underline{att}^{\varphi}$ |
| $\underline{att}^{\Delta}$ | = | deletion of entire $\underline{att}^{\varphi}$ region |
| $Int^{\Delta}$ | = | deletion of Int gene |

(2) Kellenberger, et al. (1961a)

TABLE 2

Bacterial Strains

| Strain | Relevant genotype[1] | Source and/or reference |
|---|---|---|
| W3110 | $su^-$ prototroph | J. Weigle |
| C600 | $su^+$ $att^B$ | Appleyard (1954) |
| QR93 | $su^-$ $att^\Delta$ $rec^+$ | E. Signer |
| QR94 | $su^-$ $att^\Delta$ recA | E. Signer |
| SA205 | $su^-$ $att^L$ | S. Adhya |
| SU740 | $su^-$ $att^R$ | S. Adhya |

(1) Abbreviations: $su^+$ ($su^-$) = ability (inability) to support the growth of λ sus mutants

$att^\Delta$ = deletion of the entire $att^B$ region, which is the att site for λ

$rec^+$ (recA) = ability (inability) to promote generalized recombination

$att^L$, $att^R$ = left and right prophage attachment sites

## (c) Media

All media and buffers for phage work have been described previously (Parkinson, 1968; Parkinson & Huskey, 1970).

## (d) Preparation of λ phage for DNA studies

W3110 was grown with aeration to $\sim 1 \times 10^8$/ml. at 37°C in tryptone broth supplemented with $10^{-2}$ M $MgSO_4$ and 0.2% glucose. Phage were then added at a multiplicity of 0.05 - 0.5 and aeration was continued for 5-6 hours at which time $CHCl_3$ was added to promote complete lysis. Phage yields of $5 \times 10^{10}$ to over $10^{11}$/ml. are routinely obtained with this procedure. Bacterial debris is removed by low speed centrifugation and the phage pelleted by centrifugation in the Spinco 21 rotor for 2 hours at 18,000 rev./min. The pellets are resuspended in 1/100 the initial volume with TMG buffer and the phage are then added to a saturated CsCl solution, adjusting the final density to about 1.49 g/cm$^3$. Centrifugation was for 20 to 24 hours at 30,000 rev./min. in an SW50 or SW50.6 rotor. The band of purified phage is collected by piercing the tube just below the band and collecting drops.

## (e) Heteroduplex formation

To a small test tube were added: 50 µl. of 0.2 M EDTA at pH 8.0; 0.350 ml. $H_2O$ (minus the volume of phage to be added); $5 \times 10^{10}$ particles of phage A (approximately 2.5 µg of DNA in 10 µl. of solution) and $5 \times 10^{10}$ particles of phage B. The phage DNA was extracted and denatured by the addition of 50 µl. of 1.0 N NaOH. After standing for 10 minutes at room temperature to allow complete

strand separation, the solution was neutralized by the addition of 50 µl. of 1.8 M Tris-HCl, 0.2 M Tris-base (pH ~7). Renaturation was carried out by adding 0.5 ml. of formamide (Mallinckrodt, 99%) to give a final volume of 1 ml. at pH 8.5.

Renaturation was carried out at room temperature for 1 to 2 hours, after which time the pH of the solution was 7.5 to 8.0. Further renaturation was prevented by dialysis against $10^{-2}$ M Tris, $10^{-3}$ M EDTA, pH 8 at 4°C. This preparation is then stored at 4°C; however, after several days of storage a significant fraction of the renatured molecules become cyclized. Immediately before use, decyclization is achieved by heating the heteroduplex preparation to 50°C for 5 minutes followed by an ice water quench.

### (f) Electron microscopy

Grids were prepared by the basic protein film technique (Kleinschmidt & Zahn, 1959) and stained with uranyl acetate or shadowed with platinum-palladium. Two modifications of this technique were used.

#### (i) Aqueous technique

This method is essentially that described by Davis & Davidson (1968) and in more detail by Davis, Simon & Davidson (1969). 50 µl. of a solution containing about 0.5 µg/ml. of the heteroduplex DNA preparation and 0.1 mg/ml. of cytochrome c in 0.5 M ammonium acetate, pH 7, were spread onto a clean glass slide. This solution (hyperphase) was then allowed to flow down the slide onto a clean surface of 0.25 M ammonium acetate (hypophase). A parlodian-coated grid was touched to the film about 5 mm. from the glass slide

before all of the hyperphase had run off the slide. Grids were stained with uranyl acetate as described previously (Davis & Davidson, 1968).

(ii) Formamide technique

Immediately before use the hyperphase solution is prepared containing 10 µl. of the renatured DNA preparation (0.05 µg. DNA); 10 µl. of 1 mg/ml. cytochrome c in 1 M ammonium acetate and 0.1 M Tris, pH 7.9; 40 µl. $H_2O$; 40 µl. formamide. The hypophase, also freshly prepared to eliminate pH changes which can occur in formamide solutions upon standing, contained $10^{-2}$ M ammonium acetate, $10^{-3}$ M Tris, pH 7.9, and 10% formamide or else $10^{-2}$ M Tris, pH 7.9, and 10% formamide. 50 µl. of the hyperphase were spread onto the hypophase as described in part (i) above. The film was picked up with parlodian-coated grids after one minute and the grids were stained in uranyl acetate and shadowed with platinum-palladium.

(iii) Microscopy

A Phillips EM300 electron microscope using a 50 µ objective aperture and 60 KV accelerating voltage was employed. Negatives (~3000X) were enlarged 20X or 50X on a Nikon shadowgraph and traced onto paper. Lengths were measured on these tracings with a map measurer. Occasionally the dark field mode of operation of the Phillips EM300 electron microscope was used.

# 3. RESULTS

## (a) Heteroduplex mapping of deletions and substitutions

### (i) Electron microscopy of heteroduplex DNA molecules

Accurate physical maps of deletions and substitutions in the λ chromosome have been described by Davis & Davidson (1968) and by Westmoreland, Szybalski & Ris (1969). The basic principle of this mapping method involves constructing heteroduplex DNA molecules, one strand of which contains a deletion or substitution relative to the other strand. As described in the Methods section, these heteroduplexes are formed by mixing two DNA preparations, alkaline denaturing to allow strand dissociation, and subsequently neutralizing in formamide to give gentle renaturation. The resulting DNA preparation will contain not only the two original homoduplex DNA's, but also a large proportion of heteroduplex molecules containing one strand from each of the two original molecules. If, for example, heteroduplexes were prepared with wild type DNA and DNA from a deletion mutant, the heteroduplex would contain a single-stranded loop of DNA in the wild type strand corresponding to the size and position of the DNA missing in the deletion strand. Below we describe briefly two methods for visualizing such loops.

Heteroduplex molecules mounted by the aqueous technique (see Methods section) permit accurate measurements of duplex contour lengths; however, the loops of single-stranded DNA collapse into "bushes" (MacHattie et al., 1967; Davis & Davidson, 1968). Apparently in high ionic strength and basic protein the more flexible single-stranded DNA collapses because of nonspecific base interactions. Thus in this technique the lengths of single-stranded DNA

cannot be measured even though the positions of single-stranded regions can be accurately determined because they form characteristic bush-like structures.

The formamide mounting technique (Westmoreland et al., 1969) described in the Methods section allows us to directly measure the length of single-stranded DNA in a heteroduplex molecule. Formamide is a DNA denaturing agent which under certain conditions (30 - 70% v/v, ionic strength = 0.1, 25°C) does not affect either duplex DNA or protein but does prevent single-stranded DNA from collapsing by melting out all nonspecific base interactions. In the electron microscope duplex DNA mounted with the formamide method appears essentially the same as if mounted by the aqueous method. However, single-stranded DNA appears as a flexible, nontangled filament whose contour length can be accurately measured.

Although in principle the formamide technique will yield more information, we generally used the aqueous technique for mapping deletions and substitutions. The major reasons for choosing the aqueous technique are its reliability and simplicity and the relatively short time involved in preparing and scanning grids made in this manner. In addition, the aqueous technique gives considerably fewer molecules ($\sim$ 10%) which must be discarded because of overlaps and tangles than does the formamide method. Furthermore, metal shadowing is not only unnecessary but actually obscures the single-strand bushes in heteroduplexes prepared by the aqueous mounting procedure. The formamide technique was employed whenever it was necessary to examine the length or nature of the single-stranded DNA loops in heteroduplex molecules.

(ii) Mapping principles

Figure 1 illustrates the basic principles for mapping the location and size of deletions by heteroduplex microscopy. Substitutions are also mapped in the same manner. We desire to know the fractional lengths of the two deletion endpoints from the left end of the wild type λ chromosome, the left end being defined genetically by gene A, the right end by gene R. By hybridizing the deletion mutant with λb5 DNA which contains a substitution in the right arm, we are able to mark the right arm of the heteroduplex physically and thus relate the physical map to the standard genetic map. The b5 substitution in the reference DNA also provides an internal calibration for each heteroduplex molecule so that length fluctuations between molecules, even those in the same grid, do not influence the accuracy of the measurements. Furthermore, by using an internal standard in each heteroduplex molecule, it is not necessary to include standard reference DNA molecules for length calibrations. Because of this calibration technique, we have abandoned the convention of expressing DNA lengths in microns and instead refer to sizes and locations in terms of the fractional length of $\lambda^{++}$ DNA. The values we determine are $x_1$ and $x_2$, the fractional positions of the left and right deletion or substitution endpoints measured from the left end of the λ DNA molecule.

In this mapping procedure we must first know very accurately the size and position of the b5 marker substitution. Methods for calibrating this substitution are discussed in the next section. With the b5 marker calibrated, the mapping procedure outlined in Fig. 1 is quite straightforward. We measure the three separate duplex segments (A, B and C) shown in the heteroduplex molecule in Fig. 1. The length C from the right end of the molecule to the beginning of

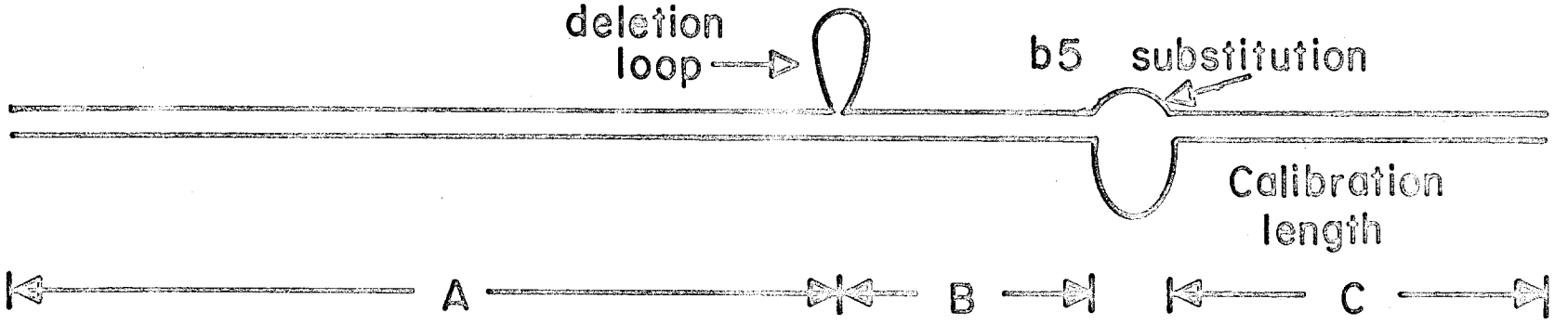
## FIGURE LEGENDS

Fig. 1. A schematic heteroduplex molecule illustrating the technique of physical mapping with the aid of an internal calibration standard. A, B and C are the measured duplex DNA lengths from the left end of the molecule to the deletion bush (A); from the deletion bush to the b5 marker substitution (B); and from the b5 substitution to the right end of the molecule (C). The fractional lengths of A and B are obtained by using length C as a standard distance, which for the b5 substitution was determined to be equivalent to 0.206 λ units. Thus,

a = fractional length of A = A/C (.206)

b = fractional length of B = B/C (.206)

The fractional position of the left deletion endpoint ($x_1$) is simply equal to a. The fractional position of the right deletion endpoint ($x_2$) is equal to 0.711, the left endpoint of the b5 substitution, minus the fractional length b. The amount of DNA deleted is therefore $x_2 - x_1$.


deletion
loop
b5 substitution
Calibration
length
A
B
C

the b5 substitution provides the necessary calibration factor in terms of $\lambda^{++}$ DNA for each molecule. The lengths A and B are converted to fractional lengths (a and b) by taking the ratio of A to C and of B to C and multiplying by the previously determined fractional length of C (designated as c). The left endpoint of the deletion ($x_1$) will equal a, the fractional length of A, which is the distance from the left end of the molecule to the beginning of the deletion. The right deletion endpoint ($x_2$) is equal to the fractional position of the beginning of the b5 substitution minus the fractional length b from the right end of the deletion to the b5 substitution. The size of the deletion is merely $x_2 - x_1$.

(iii) Calibration of reference markers

Several reference markers have been used in this work. We will describe here the methods used to calibrate the b5 marker which was most often employed. Other markers were either similarly calibrated or calibrated against b5.

The b5 substitution involves a deletion of the λ immunity region and a substitution of a different immunity region. The amount of $\lambda^{++}$ DNA missing in this mutant was determined by measuring the single-stranded lengths in heteroduplexes of $\lambda b5/\lambda^{++}$ mounted with the formamide technique. It has been found that the length ratio of single- to double-stranded DNA is not always constant; therefore ØX174 DNA which is circular and single-stranded was included on these grids as a means of calibrating single-strand lengths. This DNA is a useful calibration standard because it is circular and the possibility of measuring single-strand fragments does not exist. The size of ØX DNA relative to $\lambda^{++}$ DNA was determined by comparing the lengths of $\lambda^{++}$ DNA to those of ØX174 RFII which is circular and

also double-stranded, both DNA's being mounted on the same specimen grid. It was found that the length ratio of double-stranded ØX174 DNA to that of duplex $\lambda^{++}$ DNA was 0.112. Therefore the length ratio of single-stranded ØX DNA to a single strand of $\lambda^{++}$ DNA should also be 0.112. In this way the b5 substitution proved to be missing 0.083 of the $\lambda^{++}$ DNA which was replaced by 0.037 $\lambda$ units of DNA with the immunity of phage 21. The net deletion, therefore, must be 0.083 - 0.037 = 0.046. This value is the same as that obtained by comparing the duplex lengths of b2b5 DNA with b2 DNA. Measured against $\lambda^{++}$, the b2 deletion is missing 0.121 and the b2b5 double deletion is missing 0.166 of the wild type DNA. Thus the net amount of DNA missing in the b5 deletion is 0.166 - 0.121 = 0.045, which compares quite well with the value above.

The position of the b5 substitution was determined by measuring 60 heteroduplexes of $\lambda b5/\lambda^{++}$ mounted by the aqueous method. The length from the right end of the heteroduplexes to the beginning of the b5 substitution was measured relative to the total length of the heteroduplex. We know that the total length must be 1.000 - 0.083 so that the fractional position of the right end of the b5 substitution can be computed by simply correcting for the size of the b5 substitution. This position proves to be 0.794 ( $= x_2^{b5}$) and $x_1^{b5}$ must be 0.711 (0.794 - 0.083).

The calibration values used in this work are 0.083 which is the fractional length of $\lambda^{++}$ DNA missing in the b5 substitution and 0.206 which is the fractional distance from the right end of the $\lambda$ molecule to the right end of the b5 substitution. The values obtained by Westmoreland, Szybalski & Ris (1969) of 0.090 and 0.202, respectively, although quite similar to ours, were not used because we were unable to obtain consistent mapping data with these values.

## (b) Location and size of XOP

### (i) Use of transducing λ phages to locate XOP

Prophage integration, according to the Campbell model (Campbell, 1962), takes place through a reciprocal recombination of the phage and bacterial chromosomes at their respective attachment sites. The act of integration thus forms two hybrid attachment sites, $\underline{att}^{L}$ and $\underline{att}^{R}$, which are located at the left and right end, respectively, of the inserted prophage DNA as shown in Figure 2. The locus at which $\underline{att}^{\varphi}$ is divided during integration behaves genetically as though it were a point and therefore is designated as the cross-over point (XOP) (Parkinson, 1970). It is quite possible, however, that $\underline{att}^{\varphi}$ and $\underline{att}^{B}$ contain homologous regions at XOP so that integration cross-overs might really occur anywhere within this segment of homology. We will explore the question of homology in a later section; however, we can show by the experiments below that $\underline{att}^{\varphi}$ and $\underline{att}^{B}$ probably contain only one cross-over point each.

As may be seen from Figure 2, the cross-over region in $\underline{att}^{\varphi}$ must lie at the junction between phage and bacterial DNA in the two hybrid attachment sites $\underline{att}^{L}$ and $\underline{att}^{R}$. Thus XOP can be mapped by examining heteroduplexes between $\lambda^{++}$ and $\underline{att}^{L}$- or $\underline{att}^{R}$-containing phages. In order to recover $\underline{att}^{L}$ and $\underline{att}^{R}$ from the bacterial chromosome we have employed λ$\underline{dg}$ and λ$\underline{bio}$ transducing phages and several deletion mutants whose origins are diagrammed in Figure 3. The Campbell model specifies that transducing lambda mutants are produced by aberrant excision events in which the two excision break-points are thought to occur at random. If this is the case, then λ$\underline{dg}$ phages should contain $\underline{att}^{L}$ and λ$\underline{bio}$ phages should contain $\underline{att}^{R}$.

Fig. 2. The Campbell model of lysogeny. After entering the cell, the phage DNA circularizes by means of base pairing at the cohesive ends, pairs with the bacterial chromosome, and undergoes reciprocal recombination to insert the phage DNA into the bacterial chromosome. This recombination occurs between the att regions of the phage and bacterium which are different, to form the hybrid att sites located at the ends of the prophage.

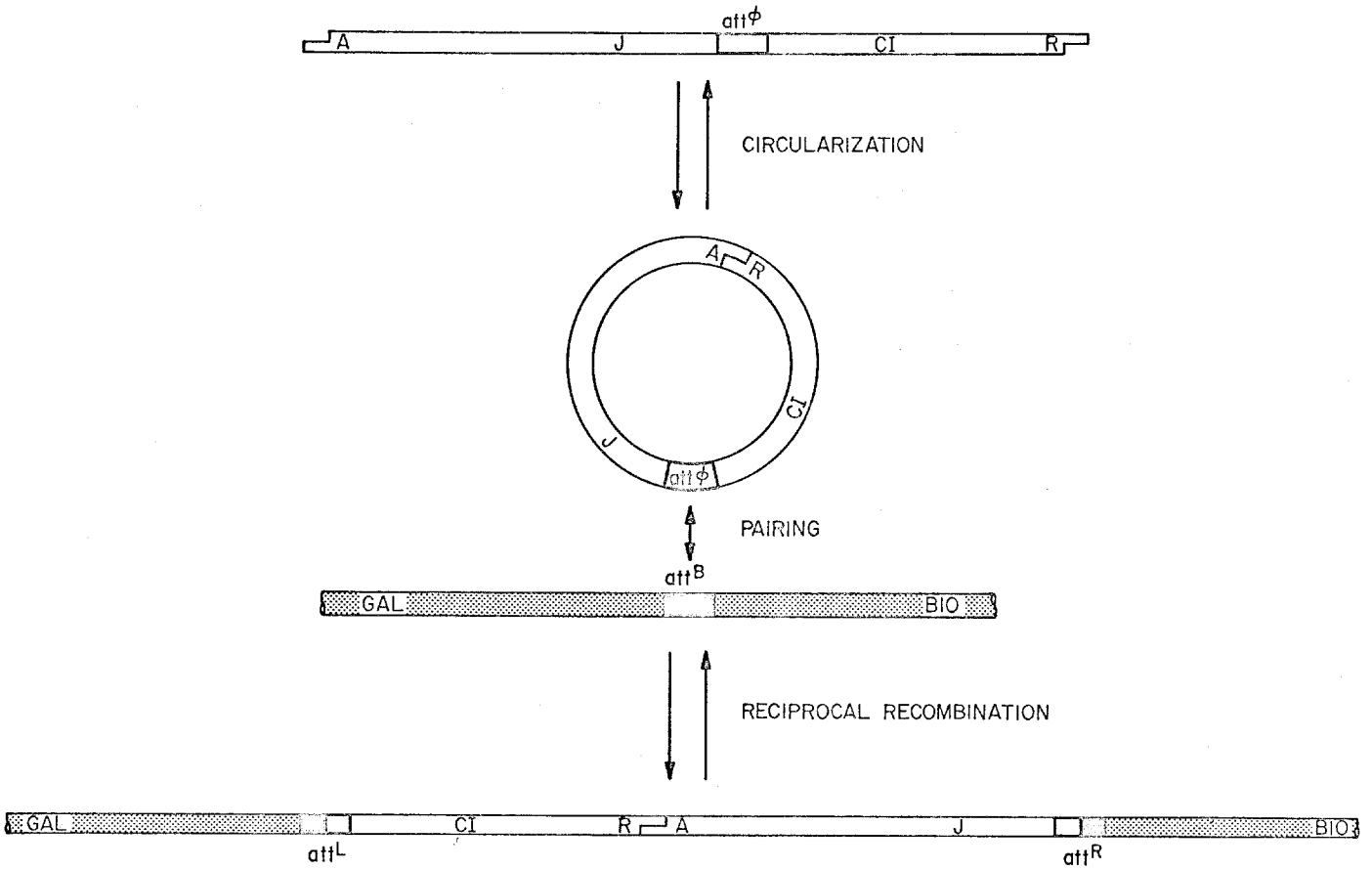

attφ
A
J
CI
R
CIRCULARIZATION
A
R
CI
J
attφ
PAIRING
attB
GAL
BIO
RECIPROCAL RECOMBINATION
GAL
CI
R
A
J
BIO
attL
attR

Fig. 3. Formation of λ phages containing $\underline{att}^{L}$ or $\underline{att}^{R}$ by aberrant prophage excision. After induction, prophage excision normally occurs by recombination of $\underline{att}^{L}$ and $\underline{att}^{R}$ to release the phage genome. Campbell (1962) proposed that rare aberrant excisions can occur in which a portion of the prophage chromosome recombines with a portion of the bacterial chromosome to produce phages containing bacterial DNA. Type one excisions are here defined as those which incorporate $\underline{att}^{L}$ DNA into the excised phage genone. Two examples of such phages are λ$\underline{dg}$ transducing mutants and deletion mutants, such as λb130, which contain $\underline{att}^{L}$, but which are not defective because no essential phage genes were lost during excision. Type two excisions produce $\underline{att}^{R}$-containing phages, exemplified by λ$\underline{bio}$ transducing mutants.

Heteroduplex molecules of λdg/λb5 and λdg/λ$^{++}$ are shown in Plate I. The large region of nonhomology in these molecules is due to the replacement of a region of phage DNA by a region of DNA containing the galactose genes from the bacterial chromosome. As expected, this region of nonhomology is in the left half of the chromosome. The right terminus of the λdg substitution should therefore correspond to the first sequence of bacterial DNA in att$^{L}$ which is not homologous to the corresponding phage DNA in att$^{\varphi}$. Thus the left-hand limit of the position of XOP is defined by the right end point of λdg substitutions. In a similar manner, the left terminus of λbio substitutions should correspond to the right-hand limit of the position of XOP.

The results given in Table 3 and summarized in Figure 4 indicate that λdg and λbio substitutions have a common end point located at 0.574 fractional length from the left end of the lambda genome. The deletion mutants, λb130 and λb189, which are shown below to contain att$^{L}$ also begin at 0.574. This end point is presumably equivalent to XOP. Since the error in these measurements corresponds to a length of no more than about 100 nucleotides, it appears most likely that XOP is in fact a fixed point or at most a region 100 nucleotides in length. The fact that λdg and λbio substitutions have a common end point at 0.574 indicates that the bacterial DNA in att$^{L}$ and att$^{R}$ is not homologous to more than 100 base pairs of the DNA in att$^{\varphi}$. Therefore, any homology between att$^{\varphi}$ and att$^{B}$ at XOP must be quite small. We will return to this problem again in a later section.

Although the data of Table 3 show that there is probably only one position in att$^{\varphi}$ at which integration can take place, it remains to be demonstrated that att$^{B}$ also has only one XOP. If, for example,

## PLATE CAPTIONS

Plate I λdg/λ heteroduplexes

a) Heteroduplex of λdgP73/λb5 mounted by the aqueous method. The arrows mark the ends of the dg bush and of the b5 marker bush. The integration cross-over point (XOP) is defined by the right terminus of the dg substitution.

b) Heteroduplex of λdgP74/$\lambda^{++}$ mounted by the formamide method. The right end of the molecule is at the right of the picture and the arrows mark the endpoints of the dg substitution. This micrograph demonstrates that dg bushes such as the one in (a) above are composed of two non-complementary single strands, one containing $\lambda^{++}$ DNA, the other containing bacterial DNA.

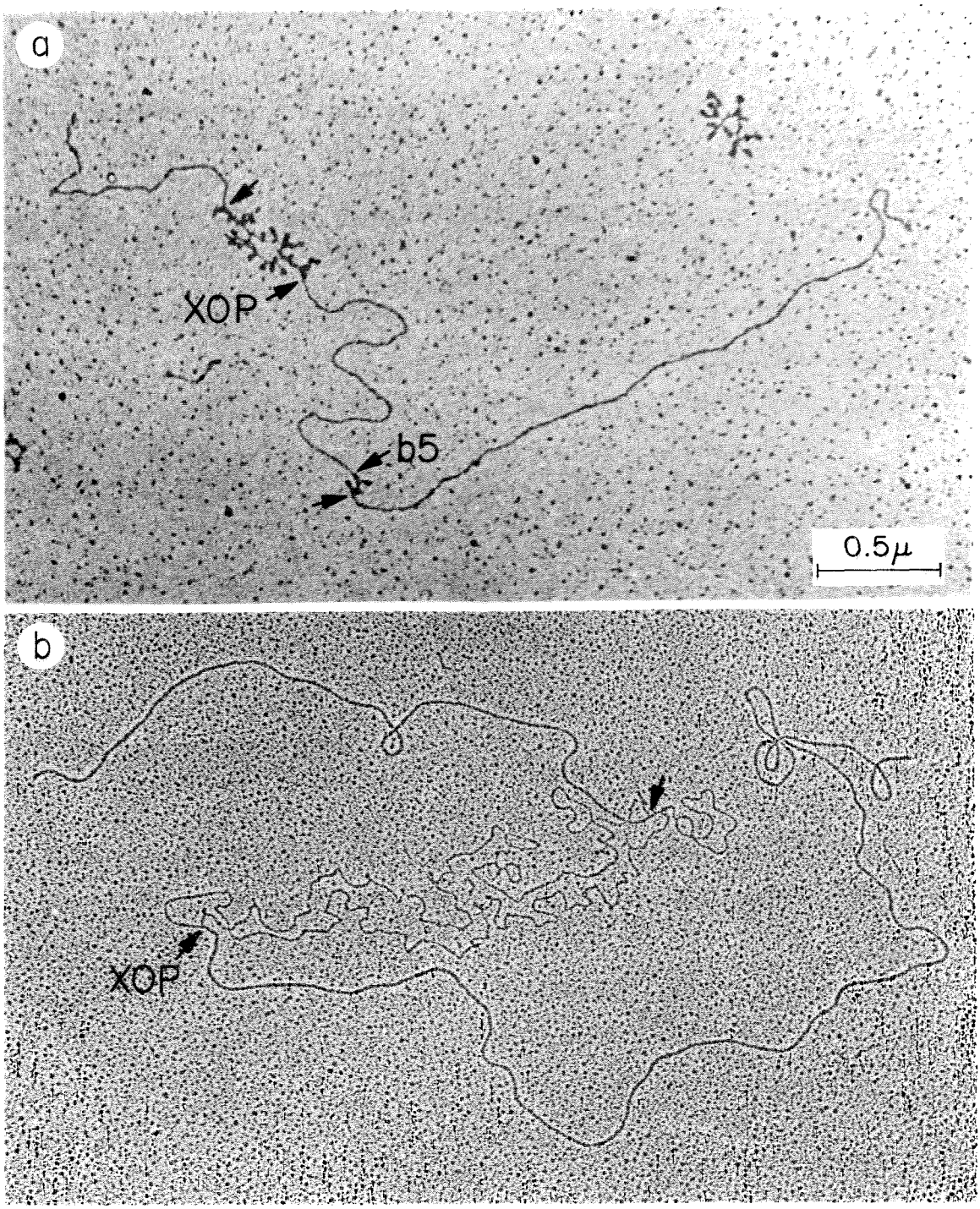

a
XOP
b5
0.5μ
b
XOP

TABLE 3

Substitution Endpoints in $\underline{att}^{L}$ and $\underline{att}^{R}$ Phages

| $\underline{att}^{L}$ phages | marker strand | number of molecules measured | Physical Position[1] $x_1$ | error | $x_2$ | error | % λ DNA deleted |
|---|---|---|---|---|---|---|---|
| λ dg P72 | b5 | 14 | .225 | ±.004 | .571 | ±.003 | 34.6 |
| λ dg P74 | b5 | 16 | .193 | ±.003 | .576 | ±.002 | 38.3 |
| λ dg P74[2] | $\lambda^{++}$ | 22 | .191 | ±.001 | ---- | ----- | 38.5 |
| λ dg P73 | b5 | 18 | .094 | ±.002 | .575 | ±.002 | 48.1 |
| λ dg P71 | b5 | 16 | .030 | ±.001 | .576 | ±.002 | 54.6 |
| λ b130[3] | b5 | 17 | .423 | ±.006 | .572 | ±.002 | 14.9 |
| λ b189[3] | b5 | 20 | .397 | ±.005 | .572 | ±.002 | 17.5 |
| SUMMARY | | 101 | Variable | | .574 | ±.001 | |

| $\underline{att}^{R}$ phages | marker strand | number of molecules measured | Physical Position[1] $x_1$ | error | $x_2$ | error | % λ DNA deleted |
|---|---|---|---|---|---|---|---|
| λ bio16A | b536 | 18 | .574 | ±.002 | .585 | ±.004 | 1.1 |
| λ bio7-20 | b536 | 20 | .573 | ±.002 | .625 | ±.004 | 5.2 |
| λ bio69 | b536 | 24 | .573 | ±.002 | .663 | ±.003 | 9.1 |
| λ bio11 | b536 | 17 | .574 | ±.002 | .677 | ±.004 | 10.3 |
| λ bio10 | b536 | 20 | .574 | ±.002 | .709 | ±.003 | 13.5 |
| SUMMARY | | 99 | .574 | ±.001 | Variable | | |

TABLE 3 (Cont'd)

(1) The physical positions of the left ($x_1$) and right ($x_2$) substitution endpoints were calculated by the procedure described in the legend of Figure 1. The error values indicate the interval which contains the mean of the population from which this sample was drawn at the 95% confidence level. The error was calculated from the equation:

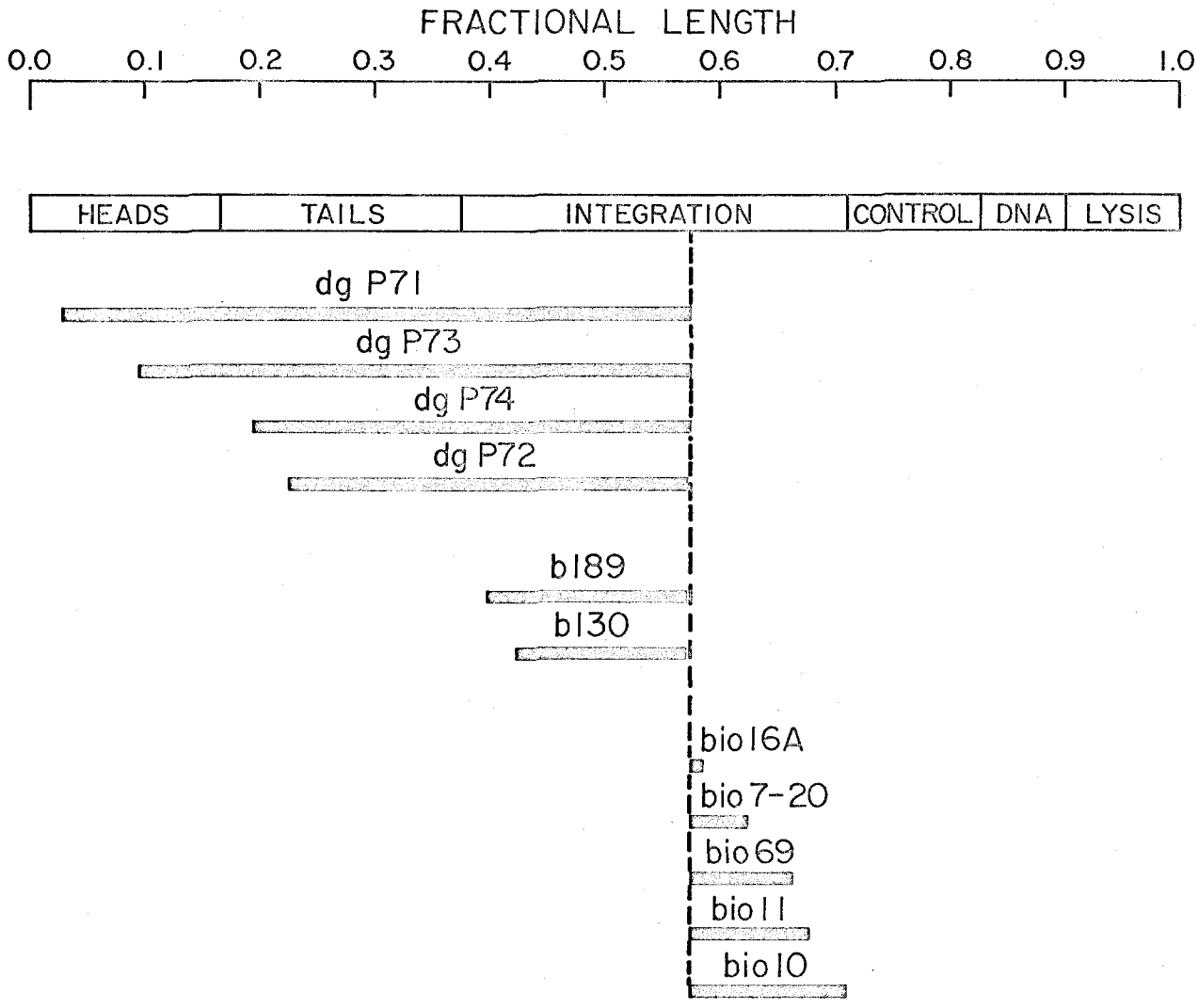
$$\text{error} = 2L_x \sqrt{1/n\,(1/L_x + 1/L_c)}$$

where $L_x$ is the average length of the duplex DNA segment measured for locating a substitution endpoint; $L_c$ is the average length of the calibrating DNA; and n is the number of measurements. This error analysis is described in more detail by Davis, et al. (1969).

(2) The left endpoint of this mutant was calculated by assuming that the substitution began at 0.574 and using the length from 0.574 to the right end as a calibration distance. This provides a check on the accuracy of the method used for the other heteroduplexes.

(3) The deletion mutants λb130 and λb189 appear to contain $att^L$ by genetic evidence (Parkinson, 1970). In a later section, λb130 is shown to contain $att^L$ by physical evidence as well.

Fig. 4. Substitution endpoints of phages containing $\underline{att}^{L}$ or $\underline{att}^{R}$. The physical locations of the substitutions in a number of phage mutants produced by aberrant excision are shown. The mapping data were taken from Table 3. All of these mutants have a common endpoint at 0.574 in the λ chromosome which is presumed to be the location of XOP. The model which describes the formation of such phage (Figure 3) predicts that one endpoint of these substitutions should not be fixed, and, in fact, this proves to be the case as shown.


FRACTIONAL LENGTH
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
HEADS
TAILS
INTEGRATION
CONTROL
DNA
LYSIS
dg P71
dg P73
dg P74
dg P72
b189
b130
bio16A
bio7-20
bio69
bio11
bio10

$\underline{att}^B$ contained two cross-over points separated by more than about 50 nucleotide pairs, we could detect this with the following experiment. A pair of λ$\underline{dg}$ or λ$\underline{bio}$ phages in which different bacterial cross-over points were used for prophage integration would form heteroduplexes containing a region of nonhomology at 0.574 due to the presence of different sequences of bacterial DNA adjacent to the phage DNA in $\underline{att}^L$ and $\underline{att}^R$. A number of λ$\underline{dg}$/λ$\underline{dg}$ and λ$\underline{bio}$/λ$\underline{bio}$ heteroduplex DNA's have been constructed between independently isolated λ$\underline{dg}$ and λ$\underline{bio}$ phages; however, no evidence has been found which suggests that $\underline{att}^B$ contains more than one XOP. Some typical examples of such heteroduplex molecules are shown in Plate II.

(ii) Use of λ deletion mutants to locate XOP

Parkinson (1970) has described several types of deletion mutants of lambda which can be used to locate XOP independently of the transducing phage method. For example, genetic mapping shows that class 2A deletions lie to the left of XOP in $\underline{att}^{\varphi}$, whereas class 3A deletions lie to the right of XOP. Therefore XOP must lie between the right end of 2A deletions and the left end of 3A deletions. When these deletions were mapped by the heteroduplex method (see Plate III), the data in Table 4 were obtained. Both the right termini of all 2A deletions and the left termini of 3A deletions map at 0.574 which is the same position found from the mapping of λ$\underline{dg}$ and λ$\underline{bio}$ substitutions. Thus XOP again is found to be located at 0.574 in the lambda chromosome.

The data of Table 4 are quite surprising in that all 2A and 3A deletions seem to begin just at XOP. Apparently these deletions have not been produced by a random process and in fact appear to be related in some way to the integration system since they have a

Plate II λdgA/λdgB and λbioA/λbioB heteroduplexes

a) Heteroduplex of λbio16A/λbio69 mounted by the formamide method and shadowed with Pt-Pd. The non-homologous region in this molecule is due to different amounts of phage and bacterial DNA's in the two λbio's. There is no detectable non-homology at XOP.

b) Heteroduplex of λbio10/λbio16A mounted as in (a) above.

c) Dark field electron micrograph of a λbio16A/λbio69 heteroduplex stained with uranyl acetate and mounted by the formamide technique. Only the region around XOP (arrow) is shown.

d) Heteroduplex of λdgP72/λdgP73 mounted as in (a) above. The segment around XOP contains no detectable non-homologous regions. Note the proximity of the molecular ends; this molecule is probably a pulled-apart circle (see Table 8).

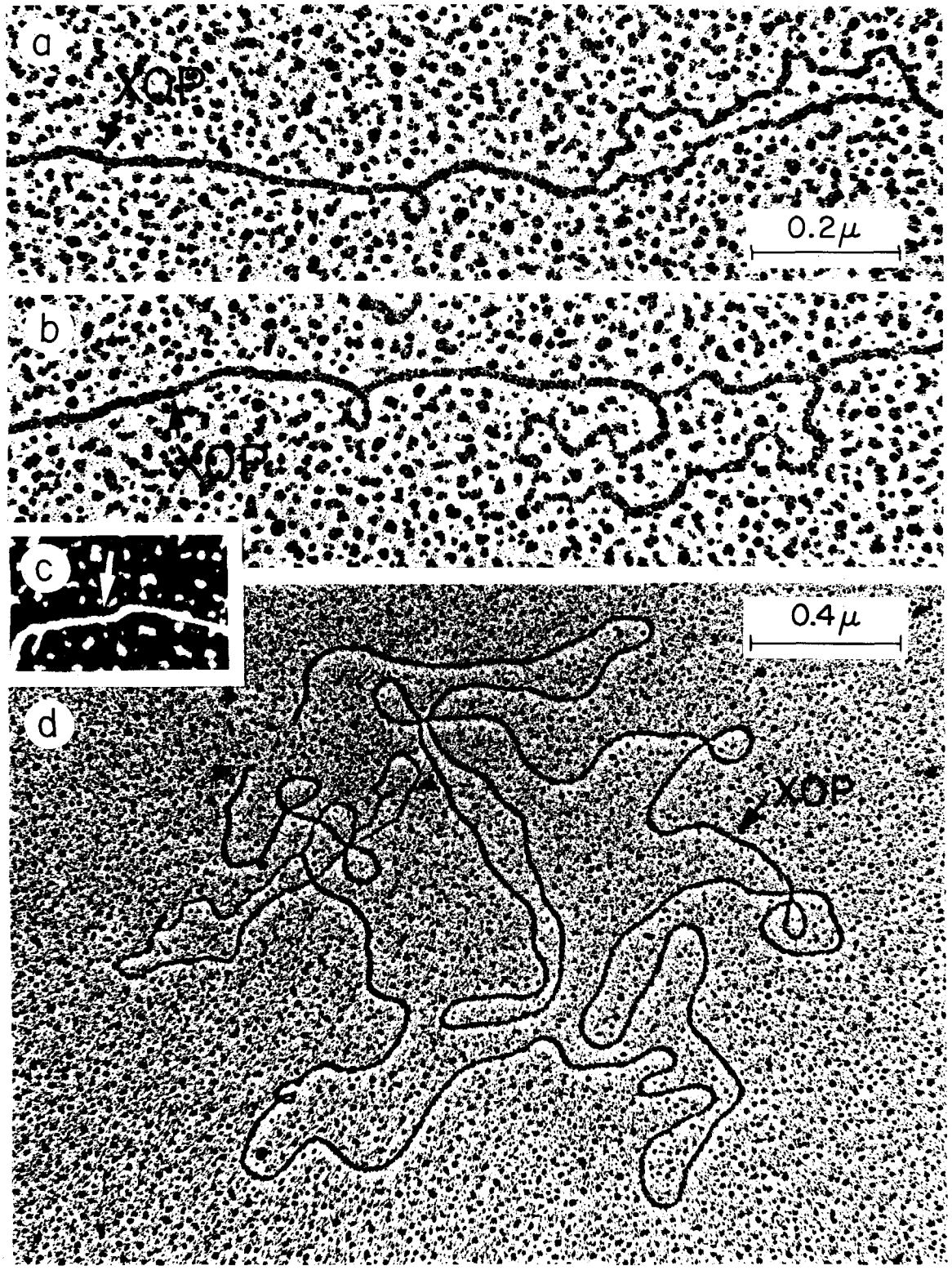

a
XOP
0.2μ
b
XOP
c
0.4μ
d
XOP

Plate III Physical mapping of deletion mutants

a) Heteroduplex of λb517/λb5 mounted by the aqueous technique. The arrows mark the position of the deletion bush and the ends of the b5 substitution bush used as a marker.

b) An enlarged view of a b5 substitution seen by the formamide mounting procedure. The arrows mark the ends of the substitution region.

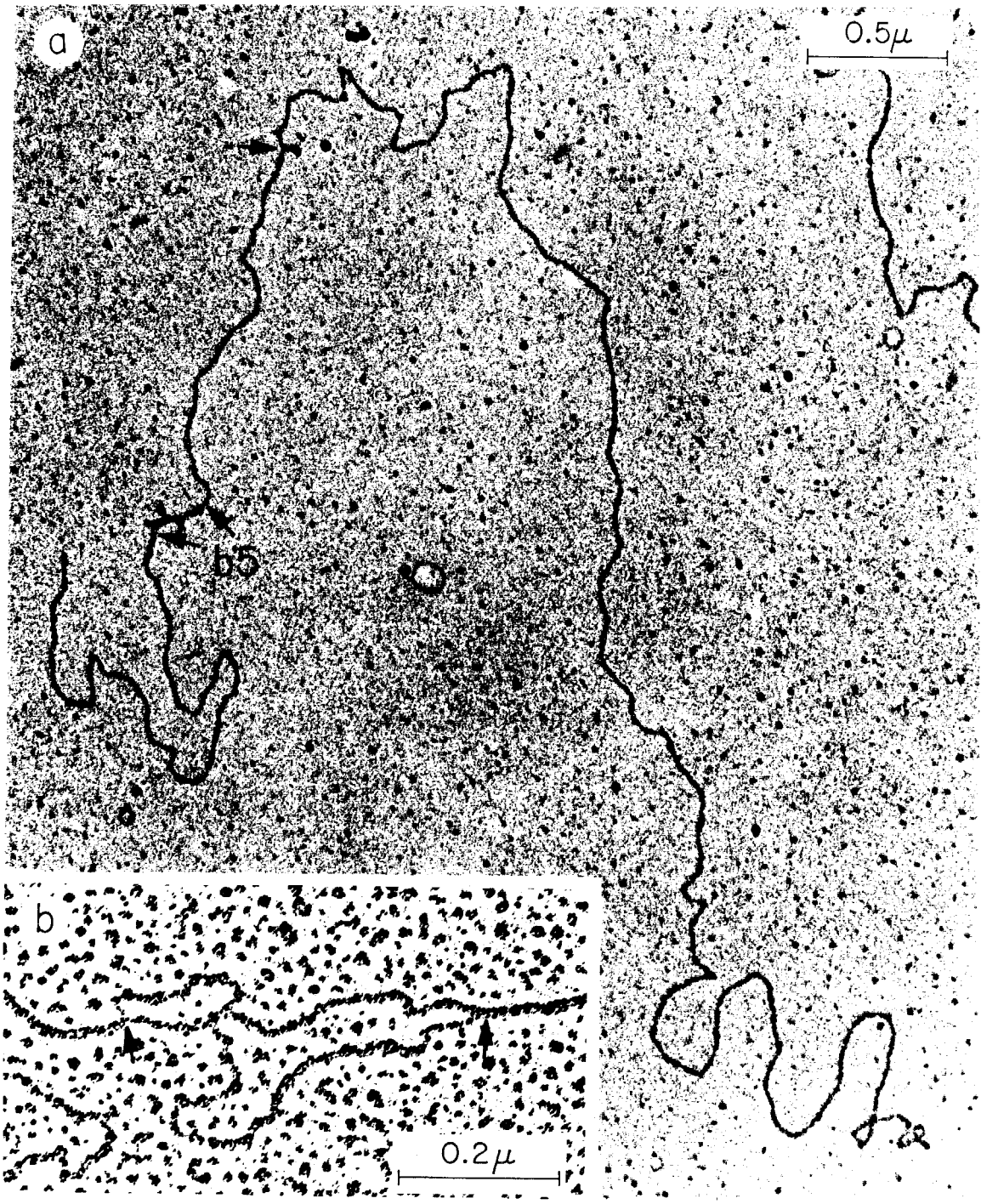

a
0.5μ
b5
b2
b
0.2μ

TABLE 4

Deletion Endpoints of Class 2A and 3A Deletion Mutants

| Class 2A mutants | marker strand | number of molecules measured | Physical Position $x_1$ | error | $x_2$ | error | % DNA deleted |
|---|---|---|---|---|---|---|---|
| b2(1) | $\lambda^{++}$ | 50 | .453 | ±.001 | .574 | ±.001 | 12.1 |
| b2 | b5 | 26 | .454 | ±.005 | .573 | ±.002 | 11.9 |
| b511 | b5 | 12 | .486 | ±.007 | .575 | ±.003 | 8.9 |
| b516 | b5 | 9 | .473 | ±.008 | .575 | ±.003 | 10.2 |
| b527 | b5 | 16 | .491 | ±.006 | .574 | ±.002 | 8.3 |
| b540 | b5 | 20 | .451 | ±.005 | .574 | ±.002 | 12.3 |
| SUMMARY | | 83 | Variable | | .574 | ±.001 | |

| Class 3A mutants | marker strand | number of molecules measured | Physical Position $x_1$ | error | $x_2$ | error | % DNA deleted |
|---|---|---|---|---|---|---|---|
| b508 | b5 | 14 | .584 | ±.008 | .613 | ±.002 | 2.9 |
| b512 | b5 | 18 | .570 | ±.007 | .614 | ±.002 | 4.4 |
| b514 | b5 | 17 | .574 | ±.007 | .614 | ±.002 | 4.0 |
| b517 | b5 | 20 | .574 | ±.007 | .616 | ±.002 | 4.2 |
| Average | | 69 | .575 | ±.004 | .614 | ±.001 | 3.9 |
| b508 | b536 | 20 | .573 | ±.000 | .610 | ±.005 | 3.7 |
| b512 | b536 | 15 | .575 | ±.001 | .614 | ±.006 | 3.9 |
| Average | | 35 | .574 | ±.001 | .612 | ±.003 | 3.8 |
| b522 | b5 | 18 | .572 | ±.007 | .643 | ±.001 | 7.1 |
| SUMMARY | | 122 | .574 | ±.001 | Variable | | |

See the footnote of Table 3. Mutants b508, 512, 514, 517 appear to be identical. These deletions are not independent isolates (Parkinson and Huskey, 1970).

(1) Data of Davis and Davidson (1968).

common end point at XOP which is presumed to be the substrate for integrase action. To test this notion spontaneous deletion mutants were isolated from int$^+$ and int$^-$ phages. Each deletion preparation was then tested for the presence of 2A deletion types by a simple spot test procedure to determine whether a functional Int gene was essential for 2A deletion formation. The data in Table 5 indicate that this is the case. In both int$^-$ preparations no class 2A deletion mutants were found, whereas the int$^+$ preparations each contained a large proportion of 2A deletion mutants. Since no spot tests are available to distinguish 3A mutants from 3B mutants, it is not possible to conclude that 3A deletions are also dependent on the Int product for their formation. This seems likely, however, since 3A deletions as well as 2A deletions begin at XOP.

Both transducing phages and deletion mutants locate XOP in att$^{\varphi}$ at 0.574 from the left end of the chromosome. A further check on this position is provided by class 3B deletion mutants which appear to lack XOP entirely (Parkinson, 1970). Mapping data for three of these deletions are presented in Table 6 and are consistent with the previous results. All of these 3B deletions are quite large and each spans the region of 0.574 where XOP is located.

### (c) Properties of REL and RER

The existence of "recognition elements" on each side of XOP in att$^{\varphi}$ has been demonstrated with att-defective deletion and substitution mutants. Class 2A deletions, for example, appear to be REL deletions since they map to the left of XOP and yet have an integration defect (Parkinson, 1970). Similarly 3A deletions are missing RER, a region to the right of XOP which is essential for normal integration

TABLE 5

Dependence of 2A Deletion Formation on Int Function

| phage | host strain | frequency of deletion mutants in original stock | number of deletions examined | fraction of deletions that were: class 1 | class 2 | class 3 |
|---|---|---|---|---|---|---|
| int$^+$ | QR93 | $6.0 \times 10^{-7}$ | 131 | 0.44 | 0.22 | 0.34 |
| int29 | QR93 | $2.8 \times 10^{-7}$ | 154 | 0.47 | 0.00 | 0.53 |
| int$^+$ | QR94 | $1.2 \times 10^{-6}$ | 156 | 0.47 | 0.24 | 0.29 |
| int29 | QR94 | $8.6 \times 10^{-7}$ | 156 | 0.86 | 0.00 | 0.14 |

Stocks of $\lambda^{++}$ and $\lambda$ int29 were grown on QR93 or QR94 by the confluent lysis method and then carried through two cycles of deletion selection with pyrophosphate inactivation as described by Parkinson and Huskey (1970). Individual plaques from the resultant deletion preparations were stabbed onto lawns of C600 and SA205, and after overnight incubation at 37°C, the turbid centers from these spots were transferred with sterile toothpicks to EMB-glucose plates spread with $\sim 10^9$ of a clear plaque mutant to test for stable lysogeny. Class 1 mutants form stable lysogens on both strains, class 3 deletions cannot stably lysogenize either strain, and class 2A mutants integrate in SA205 but not in C600 (Parkinson, 1970).

TABLE 6

Location of XOP by Class 3B Deletion Mutants

| Class 3B mutants | marker strand | number of molecules measured | Physical Position $x_1$ | error | $x_2$ | error | % DNA deleted |
|---|---|---|---|---|---|---|---|
| b221[1] | $\lambda^{++}$ | 33 | .406 | ±.003 | .629 | ±.003 | 22.3 |
| b531 | b5 | 13 | .432 | ±.006 | .605 | ±.002 | 17.3 |
| b538 | b5 | 13 | .436 | ±.006 | .599 | ±.002 | 16.3 |

See the footnote of Table 3. Mutants b531 and b538 are probably identical.

(1) Data of Davis and Davidson (1968).

behavior. From the physical positions of these and other types of deletions and the principles of deletion mapping we can locate REL and RER in the lambda chromosome and make an estimate as to their size.

By definition REL must be within the deleted region which is common to all class 2A mutants. However, class 1 deletion mutants also map to the left of XOP but are not REL-defective. Therefore REL must also lie outside of the region defined by class 1 deletions. Mapping data for a total of five different class 1 deletion mutants are presented in Table 7. These deletions all lie between 0.404 and 0.538 in the chromosome so REL must not be anywhere between these positions. Similar data for class 2A deletions has been presented in Table 4 which shows that the region between 0.491 and XOP at 0.574 is deleted in all 2A mutants that were studied. Since REL cannot lie to the left of 0.538 because of the class 1 deletions, REL must be between 0.538 and 0.574 in the chromosome. This region is approximately 2000 nucleotide pairs in length.

The problem of locating RER is more difficult because we do not yet have a class of deletion mutants on the right of XOP which are not RER-defective. Nevertheless, RER must lie somewhere between XOP at 0.574 and 0.613 which is the common region missing in both of the 3A deletions described in Table 4. This region is also about 2000 base pairs long. In Figure 5 we have summarized the use of deletion mutants to map the various components of $\underline{att}^{\varphi}$. In the Discussion section we will return to the problem of estimating the position and size of REL and RER.

TABLE 7

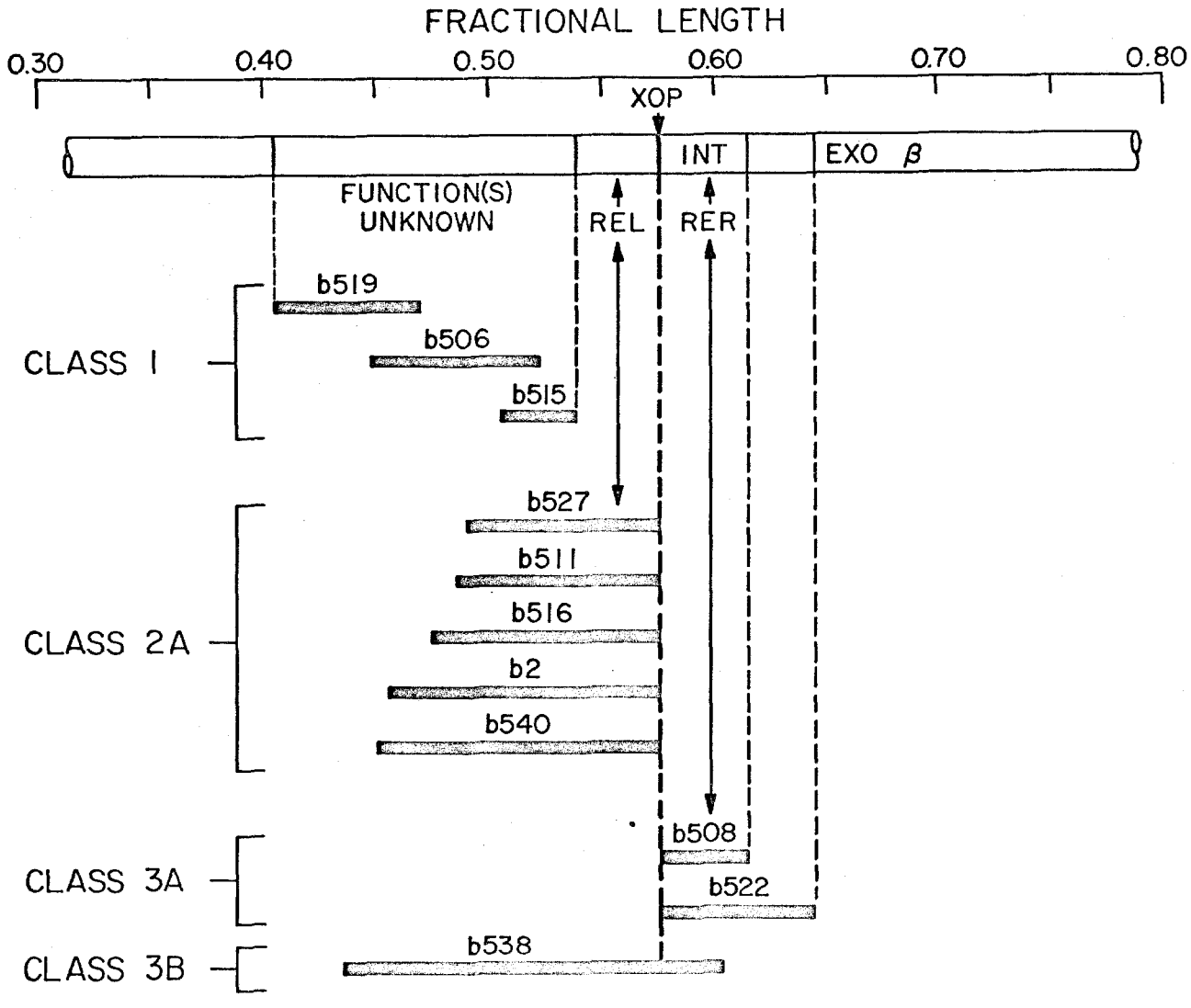
Physical Position of Class 1 Deletions

| Class 1 mutants | marker strand | number of molecules measured | Physical Position $x_1$ | error | $x_2$ | error | % DNA deleted |
|---|---|---|---|---|---|---|---|
| b501 | b5 | 12 | .465 | ±.007 | .499 | ±.004 | 3.4 |
| b502 | b5 | 12 | .465 | ±.007 | .503 | ±.004 | 3.8 |
| b504 | b5 | 13 | .467 | ±.007 | .505 | ±.004 | 3.8 |
| b509 | b5 | 11 | .470 | ±.007 | .499 | ±.004 | 2.9 |
| Average | | 48 | .467 | ±.004 | .502 | ±.002 | 3.5 |
| b506 | b5 | 10 | .447 | ±.008 | .522 | ±.004 | 7.5 |
| b510 | b5 | 14 | .399 | ±.006 | .470 | ±.004 | 7.1 |
| b519 | b5 | 13 | .407 | ±.006 | .470 | ±.004 | 6.3 |
| b520 | b5 | 14 | .405 | ±.006 | .470 | ±.004 | 6.5 |
| Average | | 41 | .404 | ±.003 | .470 | ±.002 | 6.6 |
| b515 | b5 | 11 | .504 | ±.008 | .538 | ±.003 | 3.4 |
| b536 | b5 | 20 | .428 | ±.003 | .468 | ±.003 | 4.0 |
| b536(1) | $\lambda^{++}$ | 19 | .426 | ±.003 | .463 | ±.003 | 3.7 |

See the footnote of Table 3. Mutants b501, 502, 504, 509 appear to be identical as do b510, 519, 520. These mutants are not independent isolates (Parkinson and Huskey, 1970).

(1) These values were calculated by the procedure described by Davis and Davidson (1968), and provide a check on the heteroduplex calculations for the other mutants in this table.

Fig. 5. A deletion map of the center of the λ chromosome. This map, constructed from the data of Tables 4, 6 and 7, summarizes the use of deletion mutants to study the structure of att$^{\varphi}$ and the surrounding portions of the λ genome. The class 1 deletions which have no known defects (Parkinson, 1970) define a "silent region" between 0.403 and 0.538 in the chromosome. Class 2A deletions are REL-defective and thus REL must lie between XOP and 0.538. Class 3A deletions which are RER-defective, locate RER between XOP and 0.614. The Int gene must also lie in this interval since 3A mutants are int$^{-}$. Class 3B deletions such as b538 appear to be missing the entire att region.


FRACTIONAL LENGTH
0.30
0.40
0.50
0.60
0.70
0.80
XOP
INT
EXO β
FUNCTION(S) UNKNOWN
REL
RER
b519
b506
b515
b527
b511
b516
b2
b540
b508
b522
b538
CLASS 1
CLASS 2A
CLASS 3A
CLASS 3B

## (d) A search for $\underline{att}^{\varphi}$ - $\underline{att}^{B}$ homology

### (i) Class 2B deletions contain $\underline{att}^{L}$

Parkinson (1970) has demonstrated that class 2B deletion mutants behave genetically as though they carried $\underline{att}^{L}$. We examined two of these deletion mutants ( λb130, λb189) to see whether this notion can be supported by physical evidence. The location of the non-homology between these mutants and λ wild type (Table 3) begins at XOP and extends to the left, which is consistent with both λ$\underline{dg}$ phages containing $\underline{att}^{L}$, and with simple deletions like the 2A mutants. To prove that λb130 and λb189 are not simple deletion mutants, heteroduplexes were constructed between these mutants and λ wild type and mounted for electron microscopy by the formamide procedure (Plate IVa, b). Plate IVa clearly shows that the non-homology in λb130/$\lambda^{++}$ is due to a substitution in the λb130 DNA; however, the λb189/$\lambda^{++}$ molecule in Plate IVb appears to contain a deletion loop only, with no detectable substitution in the λb189 strand.

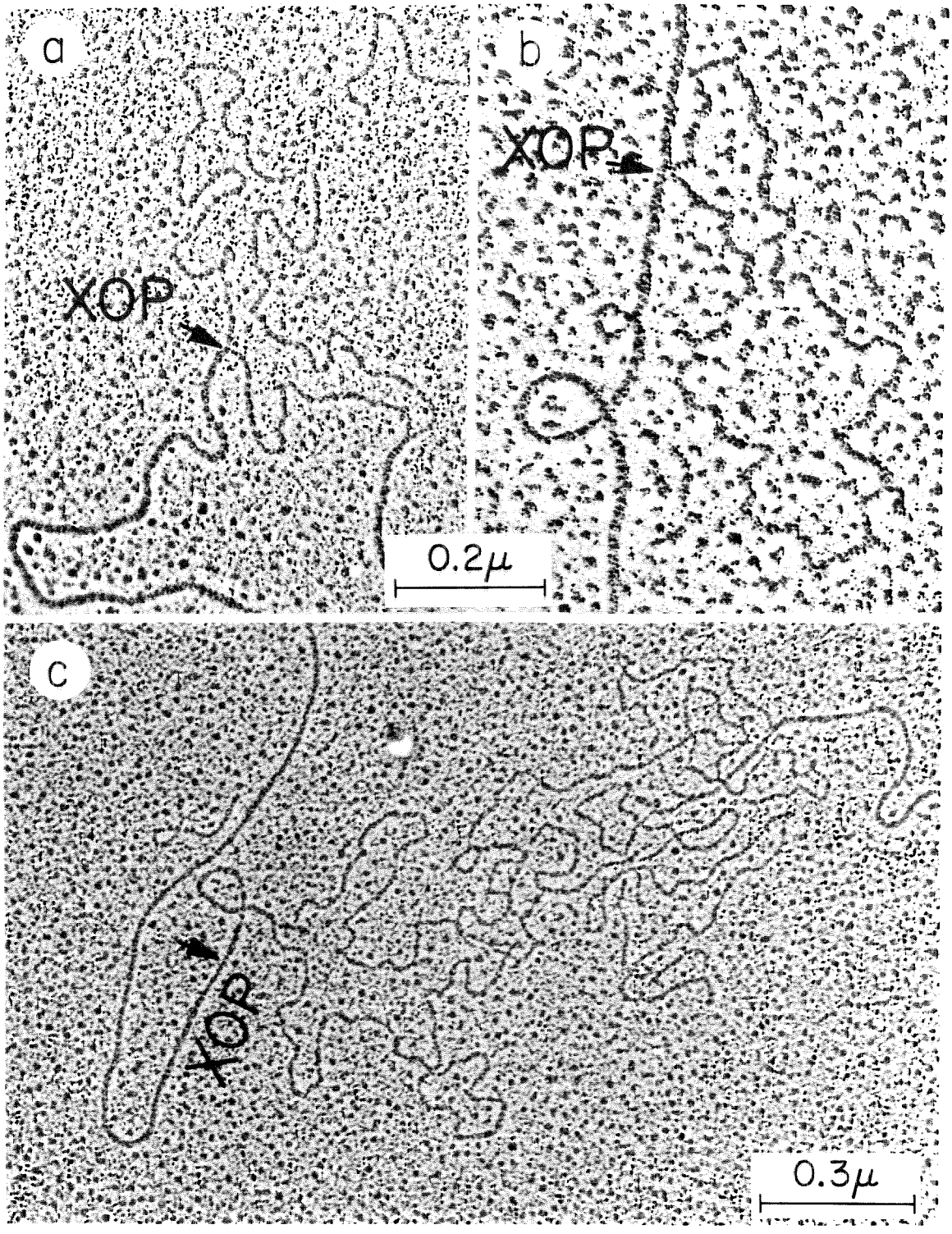
The size of the substitution in λb130 is 0.042 ± .001 λ unit. This DNA appears to be identical to the bacterial DNA at $\underline{att}^{L}$ in λ$\underline{dg}$ phages as indicated in Plate IVc which shows a λb130/λ$\underline{dg}$ heteroduplex molecule. Whereas the non-homology in λb130/$\lambda^{++}$ molecules begins at XOP, the λb130/λ$\underline{dg}$ molecules contain a duplex segment of 0.042 λ unit in length adjacent and to the left of XOP. Thus all of the substituted region in λb130 is homologous to the corresponding DNA at $\underline{att}^{L}$ in the λ$\underline{dg}$ phages. Although we cannot detect any bacterial DNA in λb189, the similarity in integration properties of λb189 and λb130 (Parkinson, 1970) supports the supposition that this mutant probably contains a small amount of bacterial DNA.

Plate IV Heteroduplex molecules containing class 2B deletions

a) Heteroduplex of λb130/$\lambda^{++}$ mounted by the formamide method. Only the region around XOP is shown to demonstrate that λb130 contains a small substitution rather than a simple deletion.

b) Heteroduplex of λb189/$\lambda^{++}$ mounted by the formamide method. The non-homology at XOP appears to be a simple deletion.

c) Heteroduplex of λb130/λdg mounted with the formamide method. A portion of λb130 DNA to the left of XOP (arrow) is homologous to the corresponding DNA from λdg. The length of λb130 DNA in (a) above that is not homologous to $\lambda^{++}$ DNA is completely homologous to λdg DNA.


a
XOP
0.2μ
b
XOP
c
XOP
0.3μ

(ii) att$^{\varphi}$ and att$^{B}$

It has been proposed that a portion of the phage and bacterial attachment sites have the same base sequence and that integration takes place by homologous crossing-over. We have already pointed out in section (b) above that there is little, if any, homology between att$^{\varphi}$ and att$^{B}$ in the vicinity of XOP. To test this notion more carefully and in a direct manner, we constructed a phage which contains att$^{B}$ to search for homology in heteroduplexes of att$^{B}$/att$^{\varphi}$. The att$^{B}$ phage was made by introducing into one chromosome the two halves of att$^{B}$ from att$^{L}$ and att$^{R}$ by crossing λb130 (att$^{L}$) by λbio69 (att$^{R}$) and selecting the λb130bio69 recombinant. The structure of this recombinant was checked by constructing the following heteroduplexes: λb130bio69/λ$^{++}$; λb130bio69/λbio16A; and λb130bio69/λdgP72 (Plate Va, b, c). The structures of these heteroduplex molecules show that λb130bio69 contains bacterial DNA on both sides of XOP and will be hereafter designated as att$^{B}$. The genetic properties of this phage are consistent with those expected of att$^{B}$ (Parkinson, 1970).

In a heteroduplex of λb130bio69/λ$^{++}$ we might be able to detect a small region of homology at XOP if, in fact, att$^{B}$ and att$^{\varphi}$ are partly homologous. However, a nucleation event is required in order to form a short duplex segment between regions of non-homology in a heteroduplex molecule. The possibility that such nucleation events cannot take place during our normal renaturation procedure was eliminated in all of the experiments described below by heating the renatured DNA in 0.1 M NaCl, 0.01 M EDTA, pH 7.3 for several hours at 45°C followed by slow cooling to 25°C.

Plate V Heteroduplex molecules containing $\underline{att}^{B}$ ($\lambda$b130$\underline{bio}$69)

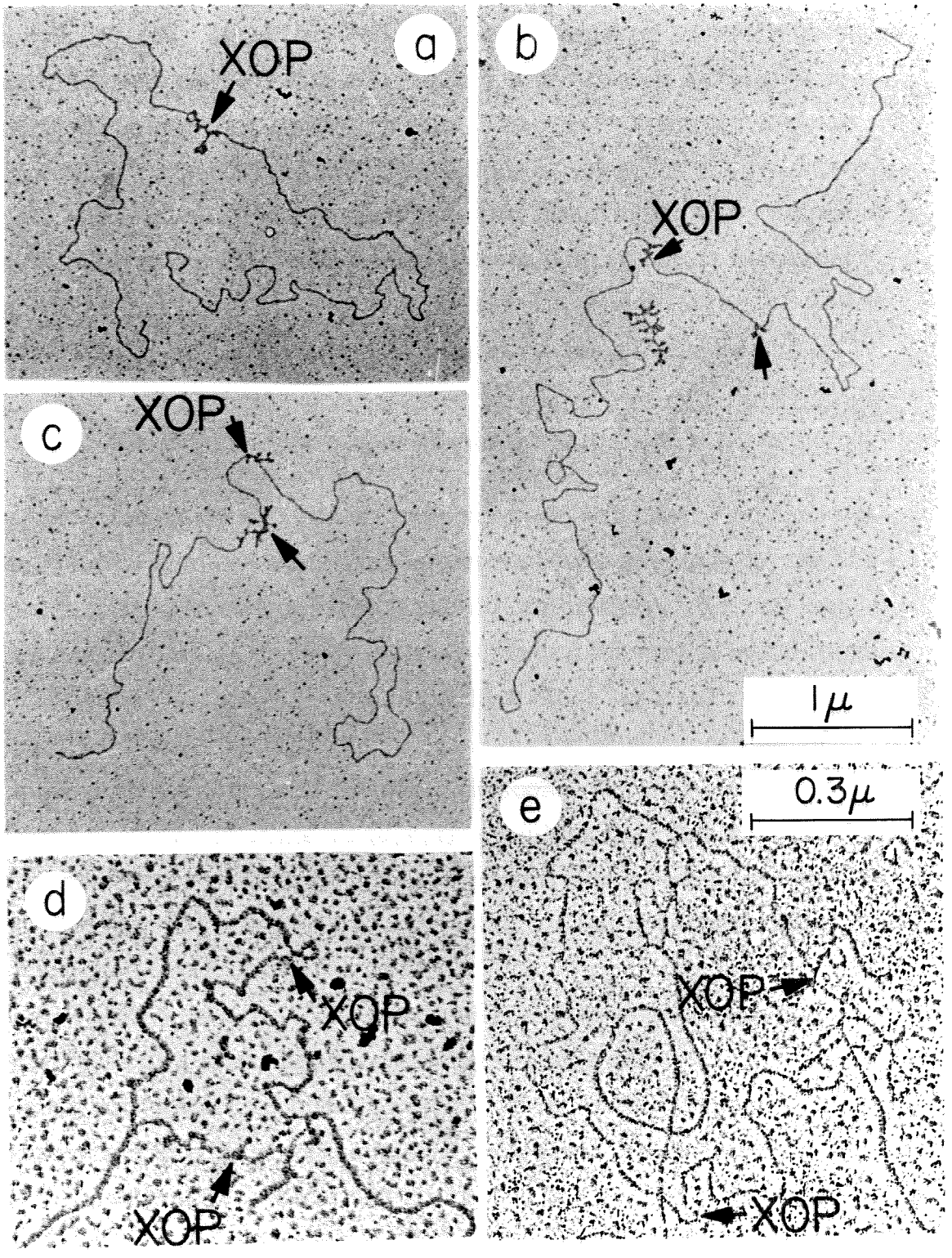
a) Heteroduplex of $\lambda$b130$\underline{bio}$69 ($\underline{att}^{B}$)/$\lambda^{++}$ ($\underline{att}^{\varphi}$) mounted by the aqueous method. The molecule contains one large bush at XOP with no detectable region of duplex DNA within the bush.

b) Heteroduplex of $\underline{att}^{B}/\underline{att}^{R}$ ($\lambda$b130$\underline{bio}$69/$\lambda\underline{bio}$16A) mounted by the aqueous method. The bush adjacent to XOP corresponds to the $\lambda$b130 deletion (Plate IVa). The other bush is due to the different substitutions in the two $\lambda\underline{bio}$'s (see Plate IIa). The region of homology between these two bushes demonstrates that $\lambda$b130$\underline{bio}$69 contains $\underline{att}^{R}$ DNA.

c) Heteroduplex of $\underline{att}^{B}/\underline{att}^{L}$ ($\lambda$b130$\underline{bio}$69/$\lambda\underline{dg}$P72) mounted with the aqueous technique. The region of non-homology adjacent to XOP is due to the $\lambda\underline{bio}$69 substitution; the other bush is from the different substitutions in $\lambda$b130 and $\lambda\underline{dg}$ (see Plate IVc). The duplex DNA between XOP and the $\underline{dg}$ bush corresponds to the length of bacterial DNA in $\lambda$b130 and proves that $\lambda$b130$\underline{bio}$69 contains $\underline{att}^{L}$ DNA.

d) Enlargement of the non-homology region in a heteroduplex of $\lambda$b130$\underline{bio}$69 ($\underline{att}^{B}$)/$\lambda$b506 ($\underline{att}^{\varphi}$) mounted by the formamide procedure using condition (g) of Table 8. There is no detectable homology between $\underline{att}^{\varphi}$ (upper strand) and $\underline{att}^{B}$ (lower strand).

e) Same as (d) except mounted in formamide with condition (a) of Table 8. The different single-strand lengths in (d) and (e) are due to differences in formamide concentration and ionic strength.


a
XOP
b
XOP
1μ
c
XOP
d
XOP
XOP
e
0.3μ
XOP
XOP

In an attempt to detect homology between $\underline{att}^{\varphi}$ and $\underline{att}^{B}$, heteroduplexes of λb130bio69/$\lambda^{++}$ were prepared as described above and mounted for electron microscopy under a wide variety of formamide and salt concentrations (Table 8). These conditions cover the range from complete melting out of random base interactions to stability of such interactions resulting in the collapse of single-stranded DNA into a bush. Even with the most favorable conditions (Plate Va), we never detected a unique region of duplex DNA in the $\underline{att}^{B}/\underline{att}^{\varphi}$ heteroduplexes (see also Plate Vd and e).

(e) Physical detection of small regions of homology

A number of deletion mutants were selected from those described in Tables 4 and 7 and heteroduplexes were constructed between pairs of deletions which should not overlap. We then searched for a short duplex segment located between the two deletion loops in these heteroduplexes to determine how close together two deletions could be before we failed to detect a region of homology between the non-overlapping ends of the two deletions. Several molecules of this type are shown in Plate VI (d, e, f). Table 9 lists the heteroduplexes that were constructed and gives a comparison of the measured length of homology with the length expected from the previously determined positions of each deletion mutant. These values compare very well, which indicates that the presence of non-homologous segments on each side of a short homologous region does not seem to cause any denaturation of that duplex segment. The λb501/λb536 heteroduplex shown in Plate VI (d, e, f) contains only about 100 base pairs of homology between the deletions and yet is perfectly stable under all formamide conditions used. Thus we feel confident that regions of

TABLE 8

Stability of Base Pairing under Various Mounting Conditions

| | Hyperphase concentration of: | | | Hypophase concentration of: | | | Presence of: | | Relative stability of random base interaction |
|---|---|---|---|---|---|---|---|---|---|
| | Formamide | $NH_4Ac$ | Tris | Formamide | $NH_4Ac$ | Tris | stable circles | "pulled-apart" circles | |
| (a) | 50% | 0.1M | 0.01M | 10% | ----- | 0.01M | - | + | unstable |
| (b) | 40% | 0.1M | 0.01M | 10% | ----- | 0.01M | - | + | unstable |
| (c) | 30% | 0.1M | 0.01M | 10% | ----- | 0.01M | - | + | unstable |
| (d) | 20% | 0.1M | 0.01M | 10% | ----- | 0.01M | - | + | partially stable |
| (e) | 10% | 0.1M | 0.01M | 10% | ----- | 0.01M | - | + | stable |
| (f) | 30% | 0.1M | 0.01M | 10% | 0.01M | 0.001M | - | + | unstable |
| (g) | 20% | 0.1M | 0.01M | 10% | 0.01M | 0.001M | - | + | partially stable |
| (h) | 25% | 0.1M | 0.01M | 5% | ----- | 0.01M | - | + | partially stable |
| (i) | 20% | 0.1M | 0.01M | 5% | ----- | 0.01M | - | + | stable |
| (j) | 50% | 0.5M | 0.01M | 10% | 0.1M | 0.01M | + | - | partially stable |
| (k) | 40% | 0.5M | 0.01M | 10% | 0.1M | 0.01M | + | - | stable |
| (l) | 0% | 0.5M | ----- | 10% | ----- | 0.01M | + | - | stable |
| (m) | 0% | 0.5M | ----- | 0% | 0.25M | ----- | + | - | stable |

Stability of base pairing in heteroduplexes of $\underline{att}^{\varphi}/\underline{att}^{B}$ was estimated by the appearance of the single-stranded regions. Self structure of this DNA appeared as short bushlike regions in the single strands. Under all of these conditions, cohesive ends of λ are stable, however, mechanical forces in the protein film during mounting often pull the joined ends apart. This type of mechanical denaturation should not be important in the stability of $\underline{att}^{\varphi}/\underline{att}^{B}$ molecules, since the region of possible homology should not be subjected to mechanical stress from the rest of the molecule.

Plate VI Detection of small regions of homology between two deletions

a), b), c) Heteroduplex of λb511/λb508 mounted by the aqueous (a) and (b) or the formamide (c) method. Any homology between these deletion mutants must correspond to the position and size of XOP. The molecules shown in (a) and (c) are typical of those found with these mounting conditions, showing that there is little or no homology between λb511 and λb508 at XOP. (b) shows an infrequently found type of molecule which shows two adjacent deletion bushes with no duplex DNA between the bushes.

d), e), f) Heteroduplex of λb501/λb536 mounted by the aqueous (d) and (e) or the formamide (f) method. A region of homology corresponding to about 100 base pairs can be detected between the deletions in these molecules. This length agrees quite well with the length expected from the position of each deletion mapped separately (see Table 9).

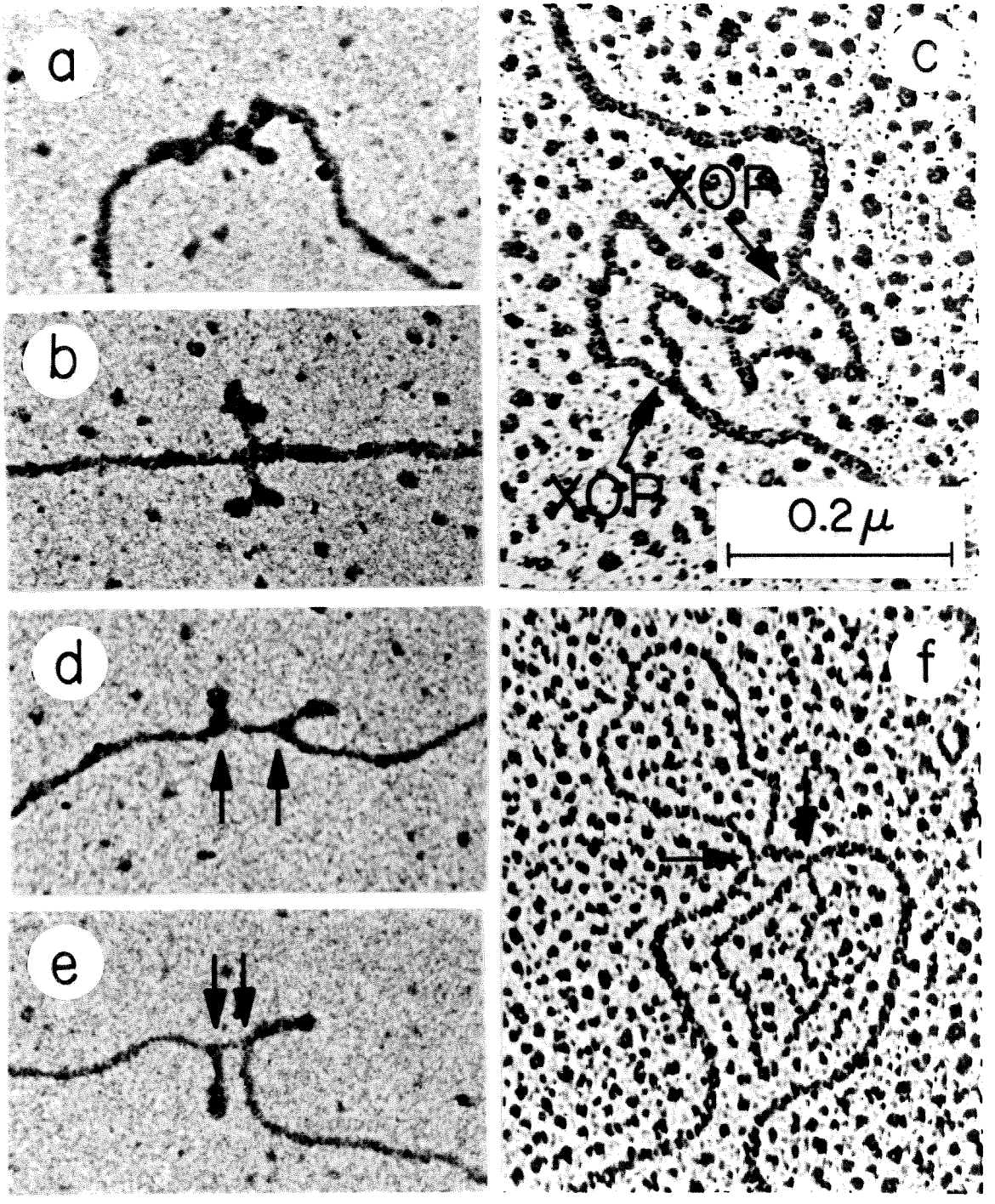

a
b
c
XOR
XOR
0.2μ
d
e
f

TABLE 9

Two Methods for Determining the Distance Between the Endpoints of Two Non-overlapping Deletions

| deletion mutants | | Length of DNA between the ends of non-overlapping deletions | |
|---|---|---|---|
| | | From data of Tables 4 and 7 | From heteroduplexes of: |
| ΔA | ΔB | $x_2^A - x_1^B$ | ΔA/ΔB |
| b536 | b501 | .002 ± .007 | .0025 ± .0002 |
| b536 | b509 | .007 ± .007 | .0027 ± .0005 |
| b536 | b511 | .023 ± .007 | .027 ± .0005 |
| b536 | b516 | .010 ± .008 | .026 ± .001 |
| b519 | b511 | .016 ± .010 | .014 ± .0005 |

Heteroduplexes between two deletion mutants were examined with the aqueous and formamide techniques, and the length of duplex DNA between the deletion loops or bushes was measured to give the values in the last column. Errors were calculated as described in the footnote of Table 3. This data indicates that b511 and b516 are probably identical.

homology even smaller than 100 nucleotides in length could be detected with the milder formamide conditions of Table 8.

The closest pair of non-overlapping deletions to be examined were 2A/3A combinations. Although each of these deletions begins at XOP, 2A and 3A mutants must not overlap since they all contain XOP and are able to recombine by Int-promoted recombination to generate wild-type recombinants (Parkinson, 1970). However, the amount of DNA between the endpoints of 2A and 3A deletions is too small to be detected by heteroduplex methods which indicates that XOP is quite small.

Heteroduplexes of λb511/λb508 do not contain any duplex segments between the deletions even when mounted under very mild formamide (Plate VIc) or aqueous conditions (Plate VIa and b). Since aqueous conditions promote complete stability of random base interactions, we should expect to detect small regions of homology under these conditions. Occasionally we found molecules like the one shown in Plate VIb that had two closely spaced deletion bushes. If we assume that the length of duplex DNA between these deletion bushes represents the maximum possible homology between 2A and 3A deletions, this homology cannot be more than 20 base pairs. More likely, the structure shown in Plate VIb is caused by the random collapse of the non-homology region into a bush which resembles two separate deletion bushes and therefore there may be no homology at all between 2A and 3A deletions at XOP.

Another homology test that we attempted involved a λb189/λb130bio69 heteroduplex molecule. If $\underline{att}^{\varphi}$ and $\underline{att}^{B}$ are actually identical at XOP, but the extent of this homology is too small to be detected with our methods, we could increase this homology with λb189 which must contain a short sequence to the left of XOP which

is identical to that in λb130. Again, however, no homology in the form of a duplex segment was detected under a variety of mounting conditions. This result suggests that the total length of XOP plus most or all of REL is still too short to be stable under these conditions.

## 4. DISCUSSION

### (a) XOP and the Int system

#### (i) Size and location of XOP

We have obtained two independent determinations of the location of XOP, the point at which the lambda chromosome is broken during integration. Analysis of both transducing phages and deletion mutants locates XOP at 0.574 in the λ genome. This value is in excellent agreement with the results of several others based on transducing phage studies. Hradecna & Szybalski (1969) reported a value of 0.573 based on heteroduplex studies of λb2 and λbio phages. Yamagishi & Skalka (1969) obtained a value of 0.57 from analyses of segments of differing base composition in $\lambda^{++}$ and λdg DNA.

The size of XOP has been estimated from studies on heteroduplexes between deletion mutants which begin at XOP and extend either to the left (Class 2A) or to the right (Class 3A). Because of the unique structure of deletion loops, which originate from a point, homology between 2A and 3A deletion mutants can be studied even in the presence of random base interactions. Two findings from these studies indicate that XOP is very small: molecules which appear to have two distinct deletion bushes are very uncommon in 2A/3A heteroduplexes; moreover, the points of origin of the two deletion bushes in

such molecules are separated by less than 20 base pairs. Since nearly all of the heteroduplex molecules do not contain two deletion bushes, the amount of homology, if any, between the XOP of 2A and 3A deletions is certainly less than 20 base pairs and perhaps less than 10 base pairs. It is difficult to estimate accurately the lower limit of detection in this case because denaturation of a small homology region could conceivably receive mechanical assistance from the forces on the rest of the molecule. This effect, however, may not be very large.

Thus XOP in $\underline{att}^{\varphi}$ is less than 20 base pairs in length which is consistent with the genetic and physical evidence that demonstrates that only one site exists in both $\underline{att}^{\varphi}$ and $\underline{att}^{B}$ for integrative cross-overs to take place. If multiple cross-over points exist, they must all be within 20 base pairs of one another and would probably not have been detected in our experiments.

(ii) Int-produced deletions

The integration and excision of the lambda prophage occurs by reciprocal recombination which appears to involve some sort of breakage and reunion event (Hoffman & Rubenstein, 1968; Ptashne, 1965). Thus one of the steps during integrative recombination is likely to involve the introduction of single- or double-stranded breaks into the phage DNA at XOP. The existence of a large number of deletions beginning at XOP and dependent on the Int gene for their formation is consistent with the notion of Int-produced nicks in $\underline{att}^{\varphi}$ at XOP. Deletions of this sort might be produced in a number of ways as a result of errors in the Int system itself or as a result of mistakes subsequent to Int action on $\underline{att}^{\varphi}$. For example, the Int

system might promote occasional unequal crossing-over between the att site of one phage and a nonhomologous region in the chromosome of another phage. Or the Int system might introduce nicks at XOP which are subsequently "chewed back" by nucleases with occasional rejoining of the chewed ends.

The most likely models of Int-dependent deletion formation seem to begin with Int-produced nicks in the attachment site at XOP. Assuming that this is the case, we can develop arguments concerning Int action based on the properties of Int-produced deletion mutants. The most striking property of the 2A and 3A deletion mutants which are produced by the Int system is that both classes of mutants contain a functional cross-over point. This suggests that XOP may actually be the junction between the recognition elements REL and RER and that one recognition element in an attachment site is sufficient to define XOP. An alternative possibility is that XOP is actually a short region similar to the cohesive ends of λDNA and generated by two staggered nicks in opposite strands of the attachment site. If this were the case, it becomes necessary to account for the fact that XOP is never lost in Int-produced deletions. One could imagine, for example, that chewing by nucleases can only occur in one direction so that XOP is never chewed away during deletion formation.

It is impossible to choose between these two alternative models of XOP and Int action on the basis of present evidence. Perhaps, however, further studies of Int-produced deletions will provide a means of studying these questions.

## (b) Structure of REL and RER

In addition to XOP, the λ att site contains two sequences of DNA, designated REL and RER, which border XOP and somehow influence the efficiency of the att site in Int-promoted recombination. Correlation of the REL-defective or RER-defective properties of deletion mutants with the physical location of these deletions in the λ chromosome allows us to set an upper limit on the size of these two recognition elements of $att^{\varphi}$. For example, REL cannot be more than 2000 nucleotides long, the limits being set on the one side by class 1 deletions which are not REL-defective, and on the other side by XOP which is present in all of the REL-defective mutants so far examined (class 2A and 2B deletion mutants).

Since all of the RER-defective mutants presently available (3A deletions, λbio transducing phages) are also $int^{-}$, we cannot as yet precisely localize RER and Int relative to XOP. However, λbio16A which is RER-defective, is only missing part of the Int gene (Manly, 1969), which suggests that RER is located between XOP and Int. Moreover, λbio16A is missing phage DNA from XOP to 0.585 in the λ chromosome. This indicates that RER is less than 500 nucleotides long.

It is quite possible that REL and RER are much smaller than the upper limits discussed above. A rather indirect line of reasoning leads us to believe that this may be the case. Our argument is based on the fact that λb189 behaves genetically as though it contained $REL^{B}$, and yet we cannot detect any bacterial DNA in this mutant. If, in fact, λb189 does contain the REL sequence from $att^{B}$, $REL^{B}$ must be less than 20 base pairs in length. Since all of the genetic and physical evidence we now have indicates that $att^{B}$ may be

structurally and functionally analogous to att$^{\varphi}$, this implies that REL$^{\varphi}$ is perhaps as small as REL$^{B}$, of the order of 20 nucleotides or less. Further analysis of the att site is necessary to demonstrate whether or not this conclusion is tenable.

### (c) Homology between att$^{\varphi}$ and att$^{B}$

It has been suggested that integrative recombination occurs through the creation of cohesive ends at the attachment site in phage and bacteria (Signer, 1968; Wu & Kaiser, 1968). This model requires only a very small amount of base sequence homology, 15 - 20 nucleotide pairs, between the two att sites. Our results, however, raise doubts as to whether there is even this much homology between att$^{\varphi}$ and att$^{B}$. Operationally, all homology between att$^{\varphi}$ and att$^{B}$ must be within the region designated as the cross-over point. We have already shown that this region appears to be smaller than 20 base pairs long and could be much smaller.

A more direct approach to this problem involved the construction of att$^{\varphi}$/att$^{B}$ heteroduplexes. No homology between these two att's was ever seen even using a wide variety of mounting conditions for electron microscopy. To estimate the sensitivity of these methods, several controls were done. Most importantly, it was demonstrated that two deletions whose endpoints are only 100 base pairs apart form two loops with a short 100 nucleotide homology region between them. This length of homology was stable even under the most denaturing conditions employed in our experiments. This suggests we could have detected considerably less homology with our methods.

Another estimate of the sensitivity of these techniques in detecting homology is that λ cohesive ends which are 20 base pairs long (Wu & Kaiser, 1968) are stable under the mild mounting con-

ditions used. This situation, however, is not strictly comparable to the case in which a small homology region is bounded by large segments of non-homology. The reason for this is that the cohered ends of λ DNA could gain additional stability through base stacking interactions at the ends of the cohered region. This effect is probably small since loss of these stacking interactions does not significantly alter the stability of cohered ends (Wang & Davidson, 1968).

There is also associated with the cyclization of λ DNA an unfavorable entropy. The entropy for forming a short duplex region in the middle of two single-stranded segments is presumably more favorable than for cyclization. Therefore, although we cannot quantitatively evaluate the stability of small homology regions bounded by large portions of non-homology, it seems likely that a region of 20 base pairs would be stable. Moreover, since the denaturing conditions used to mount the DNA were barely sufficient to melt out random base interactions, we cannot detect any homology between $\underline{att}^{\varphi}$ and $\underline{att}^{B}$ in excess of random base pairing which probably involves fewer than 10 base pairs.

Thus our results indicate that the mechanism of λ prophage integration must be a very special sort of recombination event. The enzymatic portion of this system seems to specifically recognize the short sequence of nucleotides comprising $\underline{att}^{\varphi}$ and $\underline{att}^{B}$ to bring about a reciprocal recombination between these two attachments sites. Although we do not yet know how this recombination is accomplished, it is evident that little or no base sequence homology is required for integration. Furthermore, integration cross-overs occur at only one point within the phage and bacterial $\underline{att}$ sites. In contrast, systems of generalized recombination are neither site-specific nor homology-independent. It appears, therefore, as though $\underline{att}^{\varphi}$ and the

enzymatic components of the λ integration apparatus constitute a highly specialized recombination system which may operate by quite different principles than do systems of generalized recombination.

We want to thank Drs. M. Simon, N. Davidson and R. S. Edgar for constructive criticism during the course of this work and in the preparation of this manuscript. We also would like to acknowledge the assistance of Chris Jamieson who measured several hundreds of molecules for us. This work was supported by grants to Dr. Edgar from the National Science Foundation (GB-3930) and from the National Institutes of Health (GM-6965) and by a U. S. Public Health Service grant to Dr. Davidson (GM-10991). R. W. D. was supported by U. S. Public Health Service training grant (GM-01262). J. S. P. was a predoctoral fellow under title IV of the National Defense Education Act.

## REFERENCES

Appleyard, R. K. (1954) Genetics 39, 440.

Campbell, A. (1962) Adv. Genetics 11, 101.

Davis, R. W. and Davidson, N. (1968) Proc. Nat'l. Acad. Sci. U.S. 60, 243.

Davis, R. W., Simon, M. and Davidson, N. (1969) Methods in Enzymology New York: Academic Press, in the press.

Hoffman, D. B. and Rubenstein, I. (1968) J. Mol. Biol. 35, 401.

Hradecna, Z. and Szybalski, W. (1969) Virology 38, 473.

Kaiser, A. D. (1957) Virology 3, 42.

Kayajanian, G. (1968) Virology 36, 30.

Kellenberger, G., Zichichi, M. L. and Weigle, J. (1961a) J. Mol. Biol. 3, 399.

Kellenberger, G., Zichichi, M. L. and Weigle, J. (1961b) Proc. Nat'l. Acad. Sci. U.S. 47, 869.

Kleinschmidt, A. K. and Zahn, R. K. (1959) z. Naturforsch 14B, 770.

MacHattie, L. A., Ritchie, D. A., Thomas, C. A., Jr. and Richardson, C. C. (1967) J. Mol. Biol. 23, 355.

Manly, K. F. (1969) Ph.D. thesis M.I.T. Cambridge, Mass.

Manly, K. F., Signer, E. R. and Radding, C. (1969) Virology 37, 177.

Parkinson, J. S. (1968) Genetics 59, 311.

Parkinson, J. S. (1970) J. Mol. Biol. submitted.

Parkinson, J. S. and Huskey, R. J. (1970) J. Mol. Biol. submitted.

Ptashne, M. (1965) J. Mol. Biol. 11, 90.

Signer, E. R. (1968) Ann. Rev. Microbiol. 22, 451.

Wang, J. C. and Davidson, N. (1968) Cold Spring Harbor Symp. Quant. Biol. 33, 409.

Wu, R. and Kaiser, A. D. (1968) Cold Spring Harbor Symp. Quant. Biol. 33, 729.

Westmoreland, B. C., Szybalski, W. and Ris, H. (1969) Science 163, 1343.

Yamagishi, H. and Skalka, A. (1969) Virology, in the press.

Zissler, J. (1967) Virology 31, 189.

# CHAPTER 4

## A Physical Gene Map of the Left Arm of the Lambda Chromosome

John S. Parkinson and Ronald W. Davis

The following paper has been prepared for submission to the Journal of Molecular Biology

# A PHYSICAL MAP OF THE LEFT ARM OF THE LAMBDA CHROMOSOME

It is becoming increasingly important in biophysical and biochemical studies of bacteriophage to know the physical location of genes in the phage chromosome. A number of investigators have devised various methods for constructing physical gene maps in the phages T4 (Mosig, 1966, 1968; Goldberg, 1966) and λ (Kaiser, 1962; Jordan & Meselson, 1965; Hogness, Doerfler, Egan & Black, 1966). In general, these techniques involve the measurement of the length of a DNA fragment which is capable of bearing two genetic markers. However, the accuracy of these methods is limited either by the fragment size determination (Goldberg, 1966; Kaiser, 1962; Jordan & Meselson, 1965; Hogness, Doerfler, Egan & Black, 1966) or by the determination of the genetic content of the fragments (Mosig, 1966, 1968). We describe here a simpler, more direct technique for constructing accurate physical gene maps of the λ DNA molecule. Since this method is completely independent of recombination frequency or efficiency of marker rescue from the DNA fragment, the resolution is limited only by the number of available genetic markers and the size determination of the chromosome fragment.

As a source of uniform chromosome fragments, we employ λ*dg* phages which contain large substitutions in the left arm of the λ chromosome. Since these substitutions all have a common right terminus (Parkinson & Davis, 1970), λ*dg* phages constitute a series

of overlapping deletions of the left arm. The genetic contents of these left arm fragments are determined by marker rescue tests which do not depend on the efficiency of rescue since all of the left arm material is constant in any λdg strain and thus a simple qualitative result is sufficient. The physical length of these fragments is then measured by electron microscopy of λdg/$\lambda^{++}$ heteroduplex DNA's (Parkinson & Davis, 1970; Davis, Simon & Davidson, 1970) and compared to the genetic content to construct a physical gene map.

The isolation and determination of the genetic content of λdg phages are described elsewhere (Parkinson, 1968). The construction and measurement of heteroduplex DNA's are also described elsewhere (Parkinson & Davis, 1970; Davis, Simon & Davidson, 1970), but will be summarized briefly. Heteroduplexes are formed by mixing λdg and $\lambda^{++}$ phages, lysing in NaOH to release and denature the two DNA's, and finally renaturing in formamide at neutral pH. The duplex lengths in λdg/$\lambda^{++}$ heteroduplexes are measured by electron microscopy and the fractional length of the left arm fragment is calculated using an internal calibration standard (Parkinson & Davis, 1970).

The physical and genetic endpoints of 7 different λdg strains are listed in Table 1. On the basis of these measurements, it is clear that the genetic and physical maps of the left arm are colinear. A direct comparison of the physical map and a genetic map

TABLE 1

Physical and genetic endpoints of 7 λdg phages

| λdg | Marker | No. of molecules measured | $\chi_1$ | Error |
|---|---|---|---|---|
| A914-J | b5* | 15 | .030 | ± .001 |
| C-J | b5* | 17 | .095 | ± .001 |
| F730-J | $i^{434}$ | 10 | .155 | ± .002 |
| Z405-J | $i^{434}$ | 16 | .168 | ± .0015 |
| G-J | b5* | 20 | .193 | ± .0015 |
| H706-J | b5* | 12 | .228 | ± .002 |
| K-J | $i^{434}$ | 18 | .284 | ± .002 |

$\chi_1$, the length of phage DNA in the left arm of these λdg strains, was measured using the distance from the right end of the heteroduplex molecule to the beginning of the substitution at 0.574 (Parkinson & Davis, 1970) as a calibration length. The $i^{434}$ or b5 substitution was used as a reference marker in the right arm. The error values are the limits within which the population mean should lie (Parkinson & Davis, 1970).

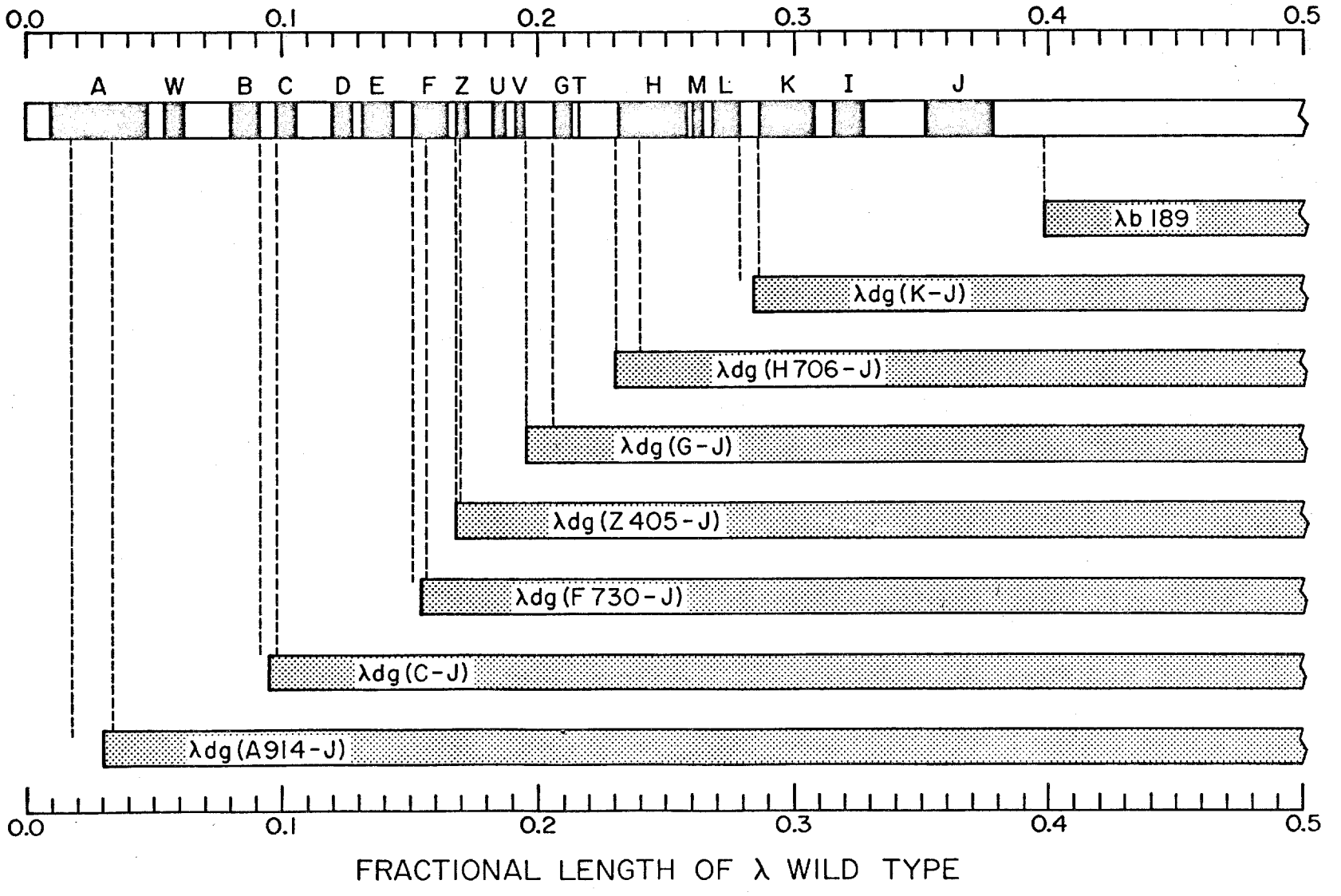
*These four λdg strains are also mapped in Parkinson & Davis (1970). The values have been recalculated using the entire right arm and central region as a calibration length as described above.

of the left arm is shown in Figure 1. The physical and genetic end-points of these 7 λdg strains can all be aligned. Thus, within the sensitivity of our measurements the frequency of recombination per unit distance is constant throughout this portion of the genome in contrast to the right arm which appears to have one region of very low recombination probability (Jordan & Meselson, 1965).

Several other features concerning the physical arrangement of genes in the left arm should be mentioned. The distance from the left end of the chromosome to the beginning of gene A is 1% of the total λ length. This segment of the genome may contain recognition sequences for the TER system which introduces the cohesive end cuts in λDNA (Gottesman & Yarmolinsky, 1968; Mousset & Thomas, 1968). Previously, this region was estimated to be about 4% in length (Jordan & Meselson, 1965). Secondly, it appears that the left arm contains few, if any, undiscovered genes. Assuming that an average gene is about 1000 nucleotides long (2% of λ length), the number of already identified genes is sufficient to fill the entire map of the left arm. Finally, the nature of the region between 0.38 and 0.4 in the genome is unclear. The right end of gene J based on this mapping data is at 0.378; however, no viable deletion mutants have ever been found to extend further to the left than 0.4. This suggests that J may extend more to the right or perhaps there is an essential gene between J and 0.4 which has not yet been identified.

In conclusion, the use of λ transducing phages to determine

FIG. 1. A comparison of genetic and physical distances in the left arm of the λ chromosome. The map at the top of this figure is the genetic map of the left arm described by Parkinson (1968). The genetic endpoints of the λdg strains, indicated by dotted lines, were fitted to the physical endpoints by projecting the genetic map onto the physical map and then expanding and contracting the genetic map to obtain maximum correspondence of the genetic and the physical endpoints as determined by direct observation. In doing this, the error in physical lengths (see Table 1) was included so that all of the genetic endpoints fall within the error limits of the physical endpoints. The limits within which the A-J genes must lie are the end of the chromosome at 0.0 and the beginning of the silent region (Parkinson & Davis, 1970) whose left end is defined by the deletion mutant, λb189 (Parkinson & Davis, 1970). The position of λ genes in this map corresponds well with physical evidence of Simon et al. (1969); however Hradecna & Szybalski (1969) have reported that the A-E distance is only 0.076 λ units whereas we obtain a value of about 0.15. We suggest that the physical or genetic mapping of the λdg strain used by these workers is incorrect.


0.0
0.1
0.2
0.3
0.4
0.5
A
W
B
C
D
E
F
Z
U
V
GT
H
M
L
K
I
J
λb 189
λdg (K-J)
λdg (H 706-J)
λdg (G-J)
λdg (Z 405-J)
λdg (F 730-J)
λdg (C-J)
λdg (A914-J)
0.0
0.1
0.2
0.3
0.4
0.5
FRACTIONAL LENGTH OF λ WILD TYPE

physical gene positions affords the most accurate method yet described for measuring distances between genes in the λ chromosome. This approach could also be employed for the right arm of the λ genome with the use of λdbio phages (Kayajanian, 1968). In conjunction with other similar mapping efforts (Parkinson & Davis, 1970; Simon, Davis & Davidson, 1969), it should be possible to construct a physical map of the entire λ chromosome which would be of great value in understanding the functional organization of the λ DNA molecule.

The authors wish to thank Drs. R. S. Edgar and N. Davidson for advice and criticism during the course of this work and in the preparation of this manuscript.

This work was supported by grants to Dr. Edgar from the National Science Foundation (GB-3930) and from the National Institutes of Health (GM-6965) and by a U.S. Public Health Service grant

(GM-10991) to Dr. Davidson. R.W.D. was supported by U.S. Public Health Service training grant (GM-01262). J.S.P. was a predoctoral fellow under Title IV of the National Defense Education Act.

Division of Biology — John S. Parkinson

and

Division of Chemistry & Chemical Engineering — Ronald W. Davis

California Institute of Technology

Pasadena, California 91109 U.S.A.

REFERENCES

Davis, R. W., Simon, M. & Davidson, N. (1970). Methods in Enzymology. New York: Academic Press, in the press.

Goldberg, E. B. (1966). Proc. Nat. Acad. Sci., Wash. 56, 1457.

Gottesman, M. E. & Yarmolinsky, M. B. (1968). Cold Spr. Harb. Symp. Quant. Biol. 33, 735.

Hogness, D. S., Doerfler, W., Egan, J. B. & Black, L. W. (1966). Cold. Spr. Harb. Symp. Quant. Biol. 31, 129.

Hradecna, Z. & Szybalski, W. (1969). Virology 38, 473.

Jordan, E. & Meselson, M. (1965). Genetics 51, 71.

Kaiser, A. D. (1962). J. Mol. Biol. 4, 275.

Kayajanian, G. (1968). Virology 36, 30.

Mosig, G. (1966). Proc. Nat. Acad. Sci., Wash. 56, 1177.

Mosig, G. (1968). Genetics 59, 137.

Mousset, S. & Thomas, R. (1968). Cold Spr. Harb. Symp. Quant. Biol. 33, 749.

Parkinson, J. S. (1968). Genetics 59, 311.

Parkinson, J. S. & Davis, R. W. (1970) J. Mol. Biol. submitted.

Simon, M., Davis, R. W. & Davidson, N. (1969) in preparation.

# CHAPTER 5

## A Physical Study by Electron Microscopy of the Terminally Repetitious, Circularly Permuted DNA from the Phage Particles of E. coli 15

Chong Sung Lee, Ronald W. Davis, and Norman Davidson

The following paper is in press in the
Journal of Molecular Biology

## Summary

Several strains of E. coli 15 harbor a prophage. Viral growth can be induced by exposing the host to mitomycin C or to uv irradiation. The coliphage 15 particles from E. coli 15 and E. coli 15 $T^-$ appear as normal phage with head and tail structure; the particles from E. coli 15 TAU are tailless. The complete particles exert a colicinogenic activity on E. coli 15 and 15 $T^-$, the tailless particles do not. No host for a productive viral infection has been found and the phage may be defective. The properties of the DNA of the virus have been studied, mainly by electron microscopy. After induction but before lysis, a closed circular DNA with a contour length of about 11.9 $\mu$ is found in the bacterium; the mature phage DNA is a linear duplex and 7.5% longer than the intracellular circular form. This suggests the hypothesis that the mature phage DNA is terminally repetitious and circularly permuted. The hypothesis was confirmed by observing that denaturation and renaturation of the mature phage DNA produce circular duplexes with two single-stranded branches corresponding to the terminal repetition. The contour length of the mature phage DNA was measured relative to $\varphi$X RFII DNA and $\lambda$ DNA; the calculated molecular weight is $27 \times 10^6$. The length of the single-stranded terminal repetition was compared to the length of $\varphi$X 174 DNA under conditions where single-stranded DNA is seen in an extended form in electron micrographs. The length of the terminal repetition is found to be 7.4% of the length of the nonrepetitious part of the coliphage 15

DNA. The number of base pairs in the terminal repetition is variable in different molecules, with a fractional standard deviation of 0.18 of the average number in the terminal repetition. A new phenomenon termed "branch migration" has been discovered in renatured circular molecules; it results in forked branches, with two emerging single strands, at the position of the terminal repetition. The distribution of branch separations between the two terminal repetitions in the population of renatured circular molecules was studied. The observed distribution suggests that there is an excluded volume effect in the renaturation of a population of circularly permuted molecules such that strands with close beginning points preferentially renature with each other. This selective renaturation and the phenomenon of branch migration both affect the distribution of branch separations; the observed distribution does not contradict the hypothesis of a random distribution of beginning points around the chromosome.

## 1. Introduction

Certain strains of E. coli derived from strain 15 lyse when treated with common inducing agents (Ryan, Fried, and Mukai, 1955; Mukai, 1960). The lysis was originally attributed to production of a colicin, but later studies showed that this lysis was due to the induction and release of bacteriophage-like particles (Sandoval, Reilly, and Tandler, 1965; Endo, Ayabe, Amako, and Takeya, 1965; Mennigmann, 1965). These particles may be defective phages; they kill certain E. coli 15 strains, but efforts to obtain a productive infection of a host have not been successful. Of the several names which have been given to these particles--colicin 15 particles, phage Ψ (Mennigmann, 1965) and coliphage 15 (Frampton and Mandel, 1968), we prefer the latter.

The morphology and colicinogenic activity of coliphage 15 have been described in the references given above. The buoyant density and melting temperature (Frampton and Mandel, 1968) of the DNA, and its degree of homology with host DNA (Cowie and Szafranski, 1965) have been reported.

The present work was initiated during a study of the small (0.67 μ) plasmid DNA molecules in E. coli 15 strain (Cozzarelli, Kelly and Kornberg, 1968) when we sought to determine whether there was any relation between the small plasmid DNA and the colicin particles. No relation has been discovered, but it was observed that after induction with mitomycin or uv light, an additional intracellular closed circular DNA of contour length 11.9 μ was formed. Later, lysis occurred and

mature phage particles were released.

This intracellular DNA is presumably a precursor of the mature linear phage DNA, but its contour length is 0.9 μ shorter. This discrepancy suggested the hypothesis that the mature DNA is terminally repetitious.

We have found that coliphage 15DNA is indeed terminally repetitious and circularly permuted. The present investigation deals primarily with properties of the DNA related to its terminal repetition and circular permutation. In part, this work is a physicochemical study of the renaturation of complementary, circularly permuted, terminally repetitious DNA strands.

Electron microscopy has been the crucial physical tool for obtaining the results reported here. Technical advances in the methodology for the study of terminal repetitions and circular permutations by electron microscopy will be described.

## 2. Materials and Methods

### (a) Bacterial strains

*E. coli* 15 wild type (WT) and *E. coli* 15 $T^-$ were gifts from Professor S. S. Cohen. *E. coli* 15 TAU and *E. coli* 15 TAU-bar were gifts from Dr. F. Funk of Professor R. L. Sinsheimer's laboratory. A gift of *E. coli* 15 TAU-bar was also generously made by Professor P. C. Hanawalt. The *E. coli* strains S/6, B/5, W3110, and C600 were from stocks of Professor R. S. Edgar and the late Dr. J. J. Weigle.

### (b) Growth media

*E. coli* 15 WT was grown in M9 buffer supplemented with 0.4% glucose, 1 mM $MgSO_4$, and 0.1 mM $CaCl_2$. M9 buffer contains 7 g of $Na_2HPO_4$, 3 g of $KH_2PO_4$, 1 g of $NH_4Cl$, and 0.5 g of NaCl per liter.

*E. coli* 15 $T^-$, TAU, and TAU-bar were grown in the above medium supplemented with 20 μg/ml thymine, 30 μg/ml uracil and cytosine, and 1/20 volume of an amino acid solution which contains 7.6 mg/ml arginine, 6 mg/ml of methionine, 2.8 mg/ml of tryptophan, and 7.5 mg/ml of proline.

Other strains of *E. coli* were usually grown in tryptone broth.

### (c) Procedures for induction by mitomycin C and ultraviolet light, and purification of the phage

Mitomycin C induction was carried out essentially as described by Endo, *et al.* (1965) and Okamoto, *et al.* (1968). Two hundred ml of a bacterial culture were grown at 37° with vigorous aeration in one of

the media described above. At a cell concentration of $2-3 \times 10^8$ cells/ml, mitomycin C (Nutritional Biochemical Company) was added to give a final concentration of 3 μg/ml. Aeration was continued for another 15 min. The cells were pelleted by centrifugation at 7000 rpm for 10 min in a Sorvall centrifuge, resuspended in 200 ml of fresh medium without mitomycin, and aeration continued. The culture became foamy and the turbidity started to decrease about 1 hr after resuspension. The total turbidity decrease was usually 20-30%. Bacterial growth was stopped when the turbidity again started to increase, by adding $CHCl_3$ and cooling to 0°.

For uv induction, 20 ml of a cell suspension ($2-3 \times 10^8$ cell/ml) in a Petri dish with a solution depth of less than 5 mm were gently swirled under a 15 watt germicidal lamp at a distance of about 40 cm. Typical growth curves are shown in Fig. 1. Exposures of 0.5, 1, and 2 min were tested; 1 min exposure was optimal. Under these conditions a λ lysogen also gives maximal lysis (Parkinson, 1969).

Unlysed bacteria and bacterial debris were removed by centrifugation in a Sorvall. Phage were then pelleted by centrifugation in a Spinco Model L centrifuge at 22,000 rpm for 3 hrs with an SW 25.2 rotor or at 30,000 rpm for 3 hrs with an SW 50.1 rotor, depending on the volume. The phage pellet was resuspended in a small volume of 0.1 M NaCl, 0.01 M Tris, 0.01 M $MgSO_4$ (pH 8). This solution was used for plaque assay and colicin tests. For the study of the DNA, the phage were further purified by adding CsCl to $\rho = 1.50$ g/ml and

buoyant banding. The phage band was collected and dialyzed against 0.1 M NaCl, 0.01 M Tris, 0.01 M $MgSO_4$ (pH 8) for storage. About $2 \times 10^{13}$ phage particles corresponding to one milligram of DNA were typically obtained from a 200-ml culture.

Colicin tests with the phage suspension were made as described by Mukai (1960).

(d) DNA extraction

Native DNA was extracted from the phage suspension after dilution to an $A_{260}$ of 0.2, either by heating for 10 min at 80° C in 0.1 M NaCl, 0.01 M EDTA (pH 7.3) or by dialysis against 8 M urea, 0.1 M NaCl, 0.01 M EDTA (pH 8.0) at 25° C, followed by dialysis against 0.1 M NaCl, 0.01 M EDTA. (All EDTA buffers were prepared by adjusting the pH of an $Na_2EDTA$ solution with NaOH.) Both procedures gave essentially complete lysis of the phage and good native DNA molecules.

(e) Isolation of closed circular DNA molecules during induction

The cells were harvested 15 min and 30 min after induction with mitomycin C. $NaN_3$ at a final concentration of 0.01 M was added to the cell suspension which was quickly chilled. The closed circular DNA was then isolated by a gentle lysis, buoyant banding in CsCl-ethidium bromide procedure (Radloff, Bauer, and Vinograd, 1967; Lee, Davidson, and Scaletti, 1968).

(f) DNA renaturation

DNA was denatured and then renatured for electron microscopy by alkaline denaturation followed by formamide renaturation. This

procedure minimizes the number of single-strand breaks introduced during the manipulations. Ten μl of phage suspension, containing 5 μg of DNA, were diluted to 0.4 ml with 0.022 M EDTA (pH 8.0). Lysis and denaturation were achieved by addition of 50 μl of 1.0 M NaOH. After 10 min the solution was neutralized by the addition of 50 μl of 1.8 M Tris-HCl, 0.2 M Tris-OH (pH ~7). Renaturation at room temperature was then achieved by the addition of 0.5 ml of formamide (Malinckrodt, 99%) and incubation in the resulting final volume of 1.0 ml for 1-10 hrs. The pH before renaturation was 8.5 and the final pH after standing was 7.5-8.0. Renaturation was stopped by dialysis against 0.1 M NaCl, 0.01 M EDTA (pH 7.3) at 5° C.

The hybridization of a mixture of coliphage 15 (WT) and coliphage 15 (TAU) DNA's was carried out similarly except that the mixture contained 5 μg of each kind of DNA. For one experiment, in which sheared wild type DNA was used with whole TAU DNA, the shearing was accomplished by forcing an alkaline DNA solution from a 0.5 ml disposable syringe through a 27 gauge needle with maximum thumb pressure 30 times. This treatment resulted in fragments of single-strand DNA with an average size of 1/7 that of the whole strand.

(g) Electron microscopy

(1) Phage. Phage particles were negatively stained with 2% phosphotungstic acid (pH 8) or 1% aqueous uranyl acetate.

(2) DNA. DNA was prepared for electron microscopy by the basic protein film technique (Kleinschmidt and Zahn, 1959). Two

variations of this technique were used:

Aqueous technique. This technique is essentially that described by Davis and Davidson (1968) and in more detail in Davis (1969). Thirty μl of a solution containing about 0.5 μg/ml of DNA and 0.1 mg/ml of cytochrome C in 0.5 M ammonium acetate (pH 7) were spread onto a clean glass slide. This solution (hyperphase) was then allowed to flow down the glass slide onto a clean surface of 0.25 M $NH_4Ac$ (hypophase). The protein-DNA mixed film was picked up on a parlodion coated grid and stained with uranyl acetate. In some cases the grids were also rotatory shadowed with platinum-palladium.

Formamide technique. When single-strand DNA is mounted by the aqueous technique it is collapsed upon itself through nonspecific intrastrand base interactions. By using a suitable concentration of formamide, both single- and double-stranded DNA can be displayed in an extended form suitable for contour length measurements and topological observations (Westmoreland, Szybalski, and Ris, 1969). One particular prescription for this procedure is the following: The hyperphase is made by mixing 10 μl of solution containing 0.05 μg of DNA with 10 μl of 1 mg/ml cytochrome C in 1 M ammonium acetate, 0.1 M Tris (pH 7.9); 40 μl of water; and 40 μl of formamide. The hypophase contains 0.01 M Tris (pH 7.9) and 10% formamide. Fifty μl of the hyperphase is spread onto the hypophase as described above. Grids were usually stained with uranyl acetate and rotatory shadowed with platinum-palladium to obtain maximal contrast.

With the formamide technique, both single- and double-stranded DNA appear as well extended threads. The latter is usually slightly thicker and more heavily contrasted. A reasonably long stretch of a DNA molecule can usually be classified with confidence as either single- or double-stranded; however, it is at present difficult to precisely locate a junction between single- and double-stranded DNA.

(3) Contour length measurements. Electron micrographs were taken on 35-mm film with a Philips EM 300 electron microscope using a 50 μ objective aperture and 60 kv accelerating voltage. Tracings were made by projecting the negative with a Nikon shadowgraph and were measured with a map measurer.

In our experience the most effective way of obtaining accurate reproducible molecular length measurements for single- and double-stranded DNA molecules is to mount the DNA molecule of unknown length and a standard DNA molecule, either single- or double-stranded to correspond with the unknown, in the same basic protein film and to photograph them simultaneously. This procedure virtually eliminates magnification errors and minimizes differential environment effects on the length of the DNA. Under favorable conditions, relative length measurements are reproducible to ± 0.2%. Even for a homogeneous DNA there is a distribution of contour lengths on a single grid due, we believe, to inherent fluctuations, at the molecular level, in the length per base pair of the DNA trapped in the film. Further information about this topic is presented in the Results section below. It is discussed in detail by Davis (1969) and by Davis, Simon, and Davidson (1969).

## 3. Results and Discussion

### (a) Properties of coliphage 15 particles

When strains of E. coli 15 wild type (WT), $T^-$, TAU, and TAU-bar were treated with uv light or mitomycin, the production of phage occurred for all cases except TAU-bar (Fig. 1) in agreement with previous results (Ryan, et al., 1955; Mukai, 1960; Mennigmann, 1965; Endo, et al., 1965, Sandoval, et al., 1965; Cozzarelli, et al., 1968). The essential results of the colicin tests, as summarized in Table 1, are that coliphage 15 (WT) and coliphage 15 ($T^-$) were colicinogenic to both strains, but not to E. coli 15 TAU and 15 TAU-bar. The particles of coliphage 15 (TAU) gave negative results with all E. coli 15 strains in this test. This defective colicinogenicity is probably due to the morphological defect described below.

Productive viral infections were not obtained with either coliphage 15 (WT) or ($T^-$) with any of the hosts of E. coli tested, including strains 15 WT, 15 $T^-$, 15 TAU, 15 TAU-bar, W3110, C600, S/6, and B/5.

The morphology of coliphage 15 (WT) as shown in the electron micrographs in Plate I (a), (b), (c), and (d) is quite consistent with the earlier observations of Mennigmann (1965) and Endo, et al. (1965). The phage has a regular hexagonal head with a diameter of 600 Å connected to a sheath of length 1000 Å and width 200 Å by a narrow neck of length 120 Å. In some cases an end plate is visible [Plate I (c)]. Many phage particles have a sheath which is quite precisely one half the full sheath length, but with a protruding core

Fig. 1. Typical growth curves for uv induction. The E. coli strains indicated in the figure were grown in 20 ml of medium as described in the text. Turbidity at 600 mμ was measured. Arrows indicate where the cell suspensions were irradiated with ultraviolet light for 1 min.

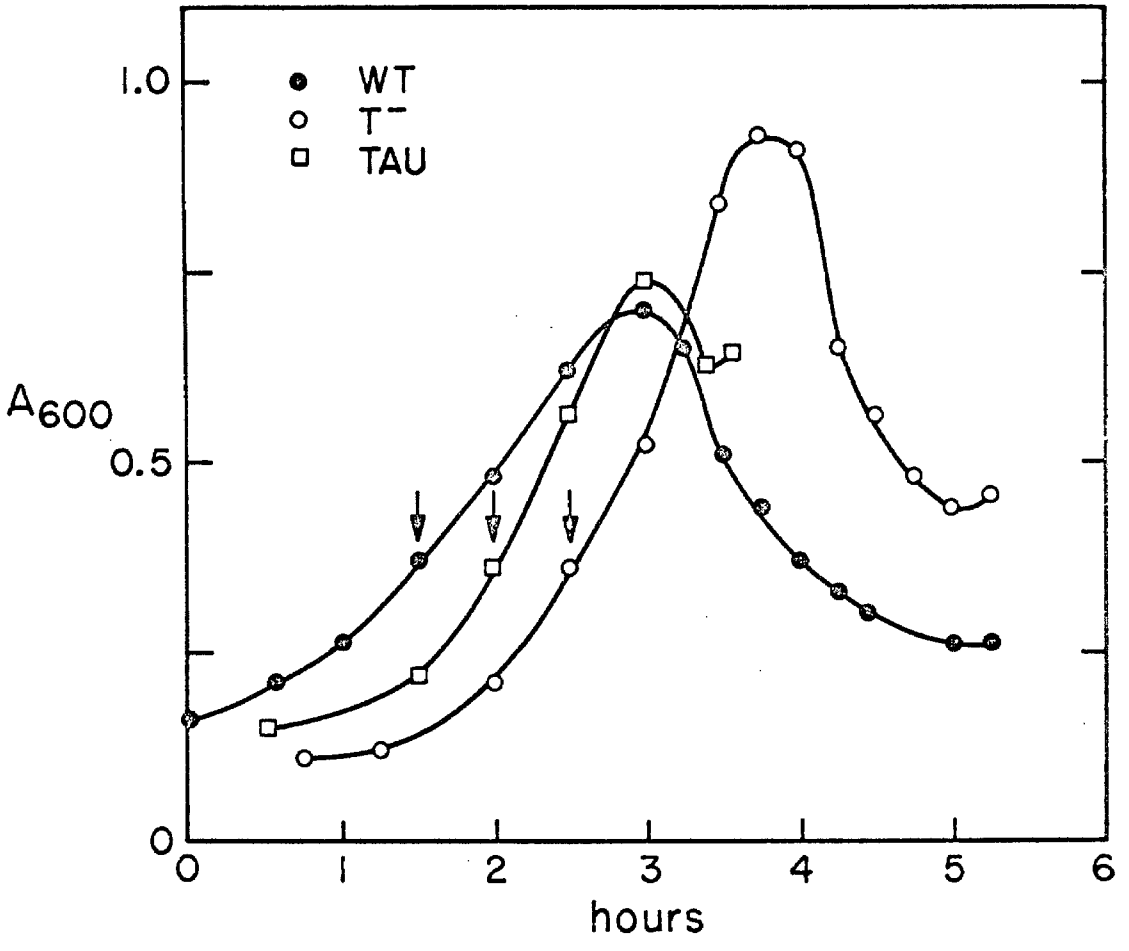

1.0
0.5
0
$A_{600}$
WT
T⁻
TAU
0
1
2
3
4
5
6
hours

TABLE 1

Properties of coliphage 15

| Strains of E. coli 15 | Inducibility (UV and mitomycin C) | Morphology of phage | Plaque assay on various E. coli strains | Colicin tests | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | WT | $T^-$ | TAU | TAU-bar | C600* |
| Wild type (WT) | + | intact phage<br>free heads<br>free tails | - | + | + | - | - | - |
| $T^-$ | + | same as above | - | + | + | - | - | - |
| TAU | + | free heads<br>no tails | - | - | - | - | - | - |
| TAU-bar | - | | | | | | | |

* In addition to C600, W3100 and B/5 were used as controls.

Plate I. Coliphage 15 (WT) and coliphage 15 (TAU) particles.

(a) An intact coliphage 15 (WT) particle from the crude lysate

(b) Common type of (WT) particle from the crude lysate showing a half sheath and core

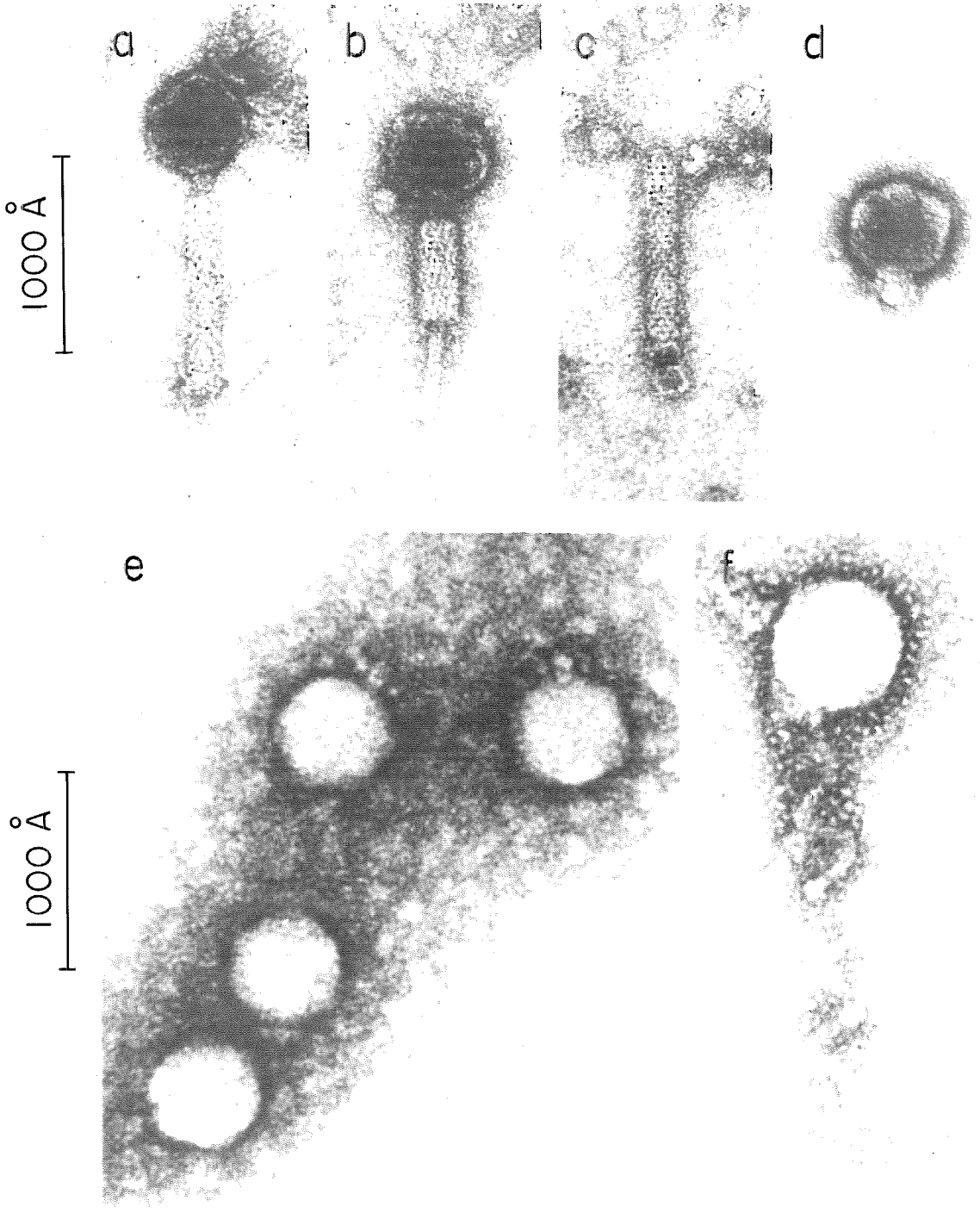
(c) Free (WT) tail showing an end plate with spikes

(d) Head of coliphage 15 (WT) after CsCl purification

(e) Head of coliphage 15 (TAU) from the crude lysate

(f) Unusual phage-like particle from 15 (TAU) crude lysate

Plates (a), (b), (c), and (d) were stained with uranyl acetate. Plates (e) and (f) were stained with phosphotungstic acid.


a
b
c
d
1000 Å
e
f
1000 Å

[Plate I (b)]. It is difficult to judge whether the sheath has become thicker (by a factor of $2^{1/2}$) as might be expected if it were a contracted full sheath. Alternatively, the full sheath may be composed of two subunits of equal length, one of which is missing in the half sheath. The lysates generally contained many more free tails than free heads as is common in phage growth. Similar observations were made with a preparation of coliphage 15 ($T^-$).

The coliphage 15 (TAU) preparations consisted entirely of tailless phage [Plate I (e)]. The dimensions and the shape of the head are very similar to those of the free head of coliphage 15 (WT) [Plate I (d)]. In a few cases spiral-like structures are seen attached to the head [Plate I (f)]. The absence of tails in the coliphage 15 (TAU) lysate could be the reason for its noncolicinogenicity, since in an earlier work it was inferred that the phage tail is the colicinogenic component (Endo et al., 1965). In spite of the difference in morphology between the TAU phage and the other strain 15 phages, they are closely related as shown by the DNA homology tests described later.

(b) DNA from mature phage particles

The native DNA molecules extracted by heat shock from the coliphage 15 (WT) and (TAU) particles appear in the electron microscope as double-stranded linear DNA [Plate II (a)] with a number average molecular length of 12.8 μ (Fig. 2).

As already mentioned, accurate ratios of contour lengths for two different DNA molecules can be measured by mounting them both on

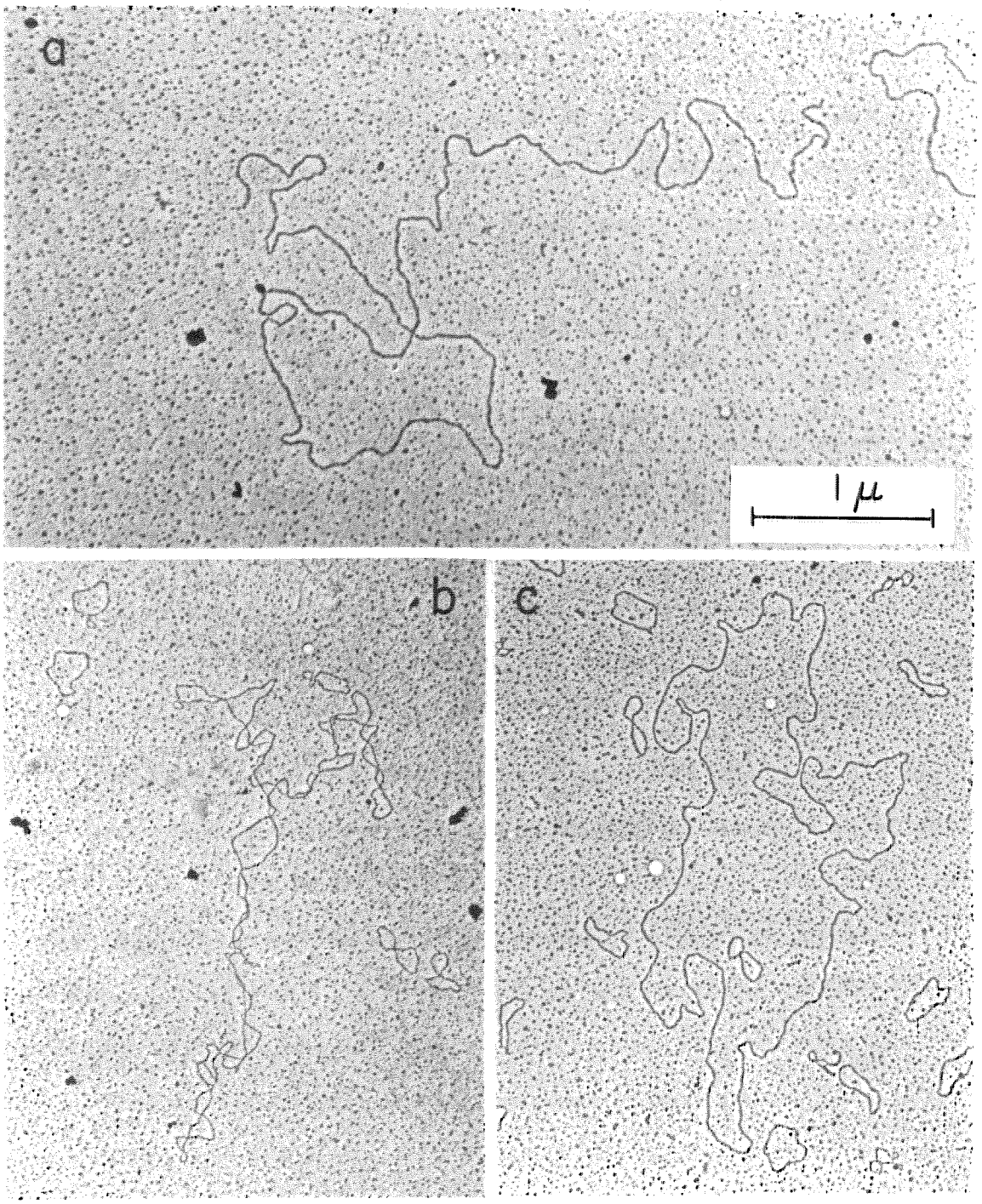
Plate II. Mature linear and *in vivo* circular coliphage 15 DNA.

(a) Mature linear coliphage 15 (TAU) DNA prepared by heat-shocking phage particles

(b) *In vivo* circular DNA observed after induction of *E. coli* TAU The large molecule is a covalently closed supercoiled coliphage 15 (TAU) DNA molecule. The small circular molecules are *E. coli* 15 minicircles.

(c) Same as (b) except that the circular coliphage 15 (TAU) DNA has an open structure.

(Mounted by the aqueous technique, stained with uranyl acetate, and shadowed with Pt-Pd.)


a
1 μ
b
c

the same grid and photographing them alternately. In this laboratory (Davis, Simon, and Davidson, 1969), we find a ratio of contour lengths of $\lambda c_{26}$ DNA and $\varphi$X 174 RF II DNA of 9.0 (± 0.1). If the molecular weight of the latter is taken as $3.4 \times 10^6$ (Sinsheimer, 1959), the molecular weight of λ DNA is calculated as $30.6(\pm 0.3) \times 10^6$. Values varying from 31 to $33 \times 10^6$ have been recommended by different authors ($31 \times 10^6$, Burgi and Hershey, 1964; $33 \times 10^6$, MacHattie and Thomas, 1964; $32 \times 10^6$, Richardson and Weiss, 1966; $33 \times 10^6$ Studier, 1965).

The ratio of contour lengths of the linear DNA of mature coliphage 15 and $\varphi$X RF II DNA was measured as 7.8(±0.1); the ratio of coliphage 15 DNA to $\lambda c_{26}$ DNA was measured as 0.875(± 0.01). For purposes of calculation, we take the molecular weights of $\varphi$X RF II and $\lambda c_{26}$ DNA as $3.4 \times 10^6$ and $31 \times 10^6$; the molecular weight of coliphage DNA is then $27.0(\pm 0.5) \times 10^6$.

No circular DNA molecules were produced by slowly cooling the linear molecules from 80° C in either 0.1 M NaCl, 0.01 M EDTA (pH 7.4) or in 2 M NaCl, 0.01 M EDTA. These are conditions which produce efficient cyclization in the lambdoid phages (Wang and Davidson, 1966a) and the N phages (Lee and Davidson, 1969). We conclude that coliphage 15 DNA does not have cohesive ends.

The complexity of the DNA (number of base pairs in one nonrepeating sequence of the DNA sample) may be determined by measuring the rate of renaturation (Wetmur and Davidson, 1968). An approximate measure of the complexity of coliphage 15 DNA was obtained

from our incidental observations of the rate of renaturation of samples prepared for electron microscopy. Under the conditions described in section (f) of Materials and Methods, λ DNA is about 50% renatured in one hour. The coliphage 15 DNA was observed to be about 64% renatured in one and one-half hours and 91% renatured in ten hours. In both cases, the degree of renaturation was scored by counting the number of two-stranded molecules and single-strand bushes in the electron microscope. Thus, the rate of renaturation is about the same as that of λ DNA. Thus, the complexity of the DNA is about equal to the number of base pairs in one mature phage molecule. The observations recorded above are probably accurate to better than a factor of two; they are certainly sufficient to eliminate the possibility that the phage DNA is a random sample of the bacterial chromosome.

The renaturation result described in a later section shows that the majority of the phage molecules do not contain any single-strand breaks.

(c) Covalently closed circular DNA molecules in vivo after induction

A culture of E. coli 15 TAU was induced with mitomycin C, samples were drawn 15 and 30 min after induction, and closed circular DNA molecules were extracted by gentle lysis and CsCl-ethidium bromide banding. Two kinds of covalently closed circular molecules were found [Plate II (b), (c)]: the E. coli minicircles of Cozzarelli et al. (1968) and a species with a contour length of 11.9 μ.

A quantitative comparison of the lengths of this circular DNA and the linear phage DNA was made by mounting a mixture of the two on the same specimen grid and photographing the two kinds of molecules alternately. For the histograms shown in Fig. 2, the difference between the number average molecular lengths of the linear and circular forms is 7.5% of the mean length of the circular form. In our experience, a difference in contour length of 0.5% can be detected by the relative length technique used here.

(d) Circular permutation and terminal repetition

If the intracellular closed circular form of DNA is indeed related to the mature linear phage DNA, the fact that the circular molecule is 7.5% shorter could be due to a terminal repetition of this length in the mature linear phage DNA. Several other phage DNA's which have a relatively large terminal repetition of this magnitude are also circularly permuted (T2, MacHattie, Ritchie, Thomas, and Richardson, 1967; P22, Rhoades, MacHattie, and Thomas, 1968). Mature P1 DNA is 12% longer than the closed circular prophage DNA, and is circularly permuted and terminally redundant (Ikeda and Tomizawa, 1968). We therefore searched for circular permutations and terminal repetitions in the DNA by the denaturation-renaturation method of Thomas and MacHattie (1964). The principles of this method are illustrated in Fig. 3.

Denaturation followed by renaturation of an ensemble of circularly permuted molecules labeled Form I in Fig. 3 will initially

Fig. 2. Histogram of molecular lengths of mature linear phage DNA and in vivo circular DNA. The in vivo circular DNA and the mature linear phage DNA were mixed and mounted on the same grid. Linear and circular molecules were alternately photographed without refocussing. Thus, an accurate, direct comparison of the relative lengths of the two species was possible. The absolute lengths of the two species were obtained using a diffraction grating for calibration and are, thus, only approximate. Since the relative length of the two species is the most significant quantity, each distribution is plotted relative to the mean length of the circular DNA. The unshaded area is for 59 linear molecules. The shaded area is for 24 circular molecules. $\underline{L}_i$ is contour length. The number average molecular length of the linear molecules, $\langle \underline{L} \rangle_{\underline{n}}^{\underline{l}}$, is 12.8 $\mu$, with a $r$ of 2.7%. The number average molecular length of the circular molecules is $\langle \underline{L} \rangle_{\underline{n}}^{\underline{c}} = 11.9\ \mu$, with a $r$ of 2.2%. The difference in mean lengths, 0.90 $\mu$, divided by the mean length of the circular molecules is 0.075; this quantity is measurable with a standard error of 0.005.

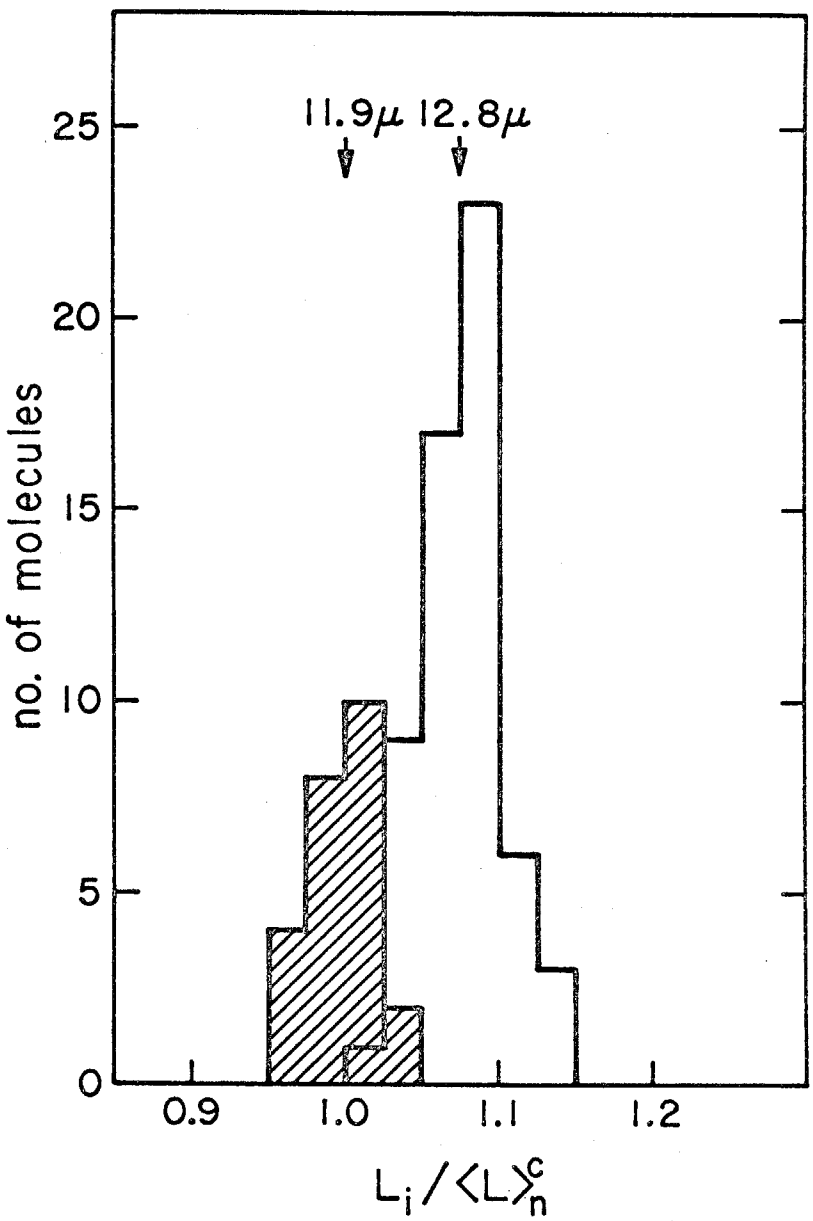
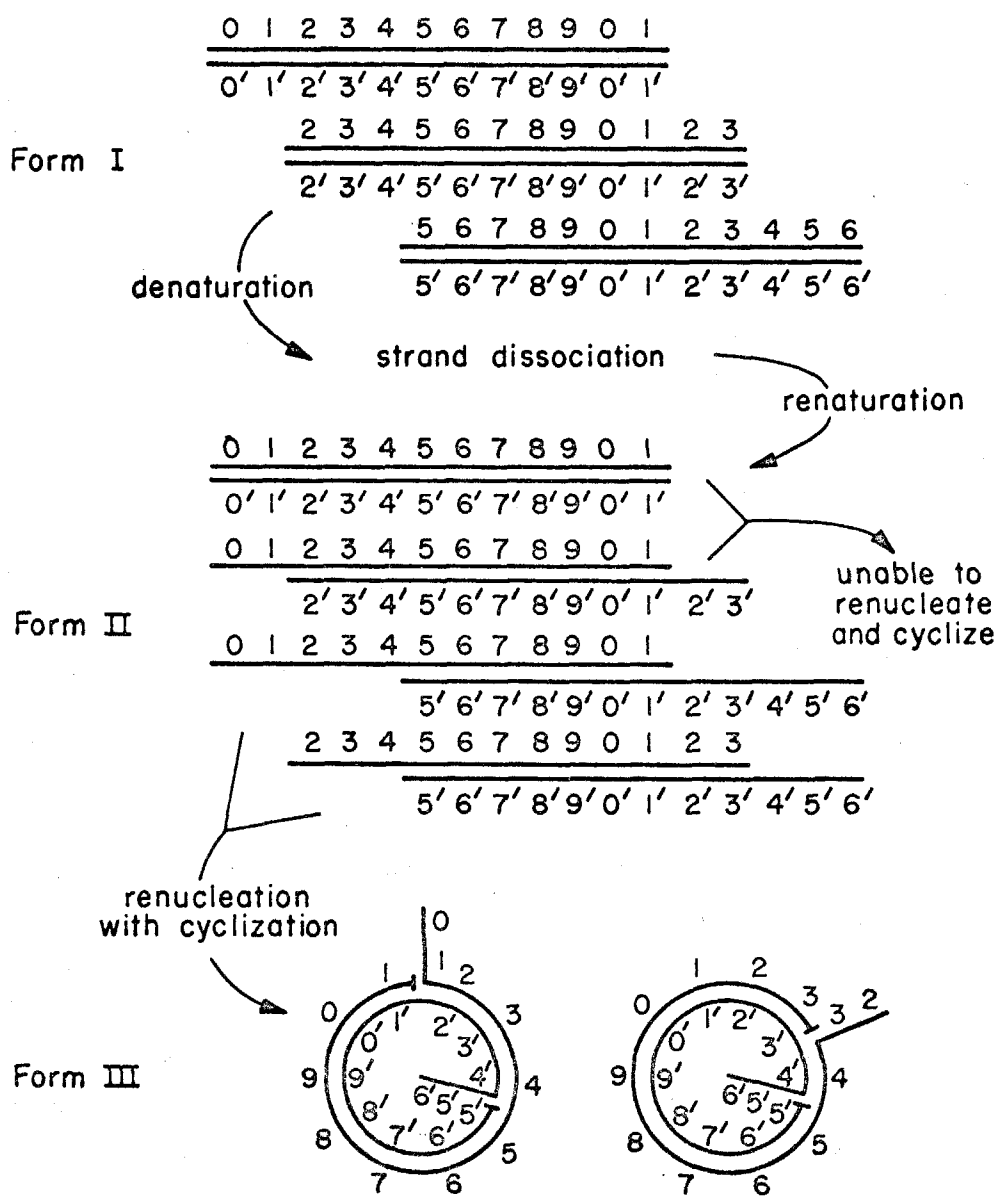

11.9μ
12.8μ
25
20
15
10
5
0
no. of molecules
0.9
1.0
1.1
1.2
$L_i / \langle L \rangle_n^c$

Fig. 3. Experimental method for studying circularly permuted and terminally repetitious DNA (slightly modified from Fig. 3 of MacHattie, et al., 1967).


Form I
0 1 2 3 4 5 6 7 8 9 0 1
0′ 1′ 2′ 3′ 4′ 5′ 6′ 7′ 8′ 9′ 0′ 1′
2 3 4 5 6 7 8 9 0 1 2 3
2′ 3′ 4′ 5′ 6′ 7′ 8′ 9′ 0′ 1′ 2′ 3′
5 6 7 8 9 0 1 2 3 4 5 6
5′ 6′ 7′ 8′ 9′ 0′ 1′ 2′ 3′ 4′ 5′ 6′
denaturation
strand dissociation
renaturation
Form II
unable to renucleate and cyclize
renucleation with cyclization
Form III

lead to the collection of linear molecules labeled Form II in the figure. Renaturation of two strands which begin at different points initially leads to a linear duplex with single-strand ends. In some of the Form II molecules the single-strand ends are mutually complementary and the molecules can cyclize to give Form III molecules. If terminal repetition occurs, the resulting circular molecules have two single-strand branches consisting of repetitious pieces. The length of the branches is the length of the terminal repetition. The separation of the branches is related to the distance between the starting points of the two strands in the renatured duplex. Linear molecules with single-strand ends also occur (Form II); some of these molecules have short single-strand ends of length less than the terminal repetition and are unable to cyclize; others have longer single-strand ends with mutually complementary regions but have failed to cyclize for kinetic reasons.

Single-strand DNA appears as a bush when mounted in the basic protein film from an aqueous medium; it appears spread out as a filament when mounted at a suitable formamide concentration, as discussed in Materials and Methods. When coliphage 15 DNA molecules are denatured by alkali, renatured in formamide, and mounted from an aqueous medium for electron microscopy, many circular molecules with single-strand bushes are seen. Plate III (a and b) gives examples of such molecules. Plates IV and V have several examples of circular molecules mounted from a formamide

Plate III. Renatured coliphage 15 DNA.

(a) Renatured circular coliphage 15 (TAU) DNA molecule showing two single-stranded repetitious branches. This molecule corresponds to Form III of Fig. 3.

(b) Same as (a) except the bushes are closely spaced showing a different permutation

(c) Renatured linear coliphage 15 (TAU) DNA molecule showing two single-stranded ends (end bushes). In the upper right corner is a whole single strand of coliphage 15 (TAU) DNA. Note that the end bushes are of equal size. This molecule corresponds to Form II of Fig. 3.

(d) Renatured sheared coliphage 15 (WT) DNA. Note the short duplex length and the absence of any large end bush. This molecule is designated as a fragment/fragment homoduplex. [Section (g) in Results].

(e) Renatured whole coliphage 15 (TAU) DNA with sheared coliphage 15 DNA. This molecule is designated as a whole/fragment heteroduplex. Note the short duplex length and the difference in size of the end bushes. The right end bush is almost as large as a whole single strand of DNA [upper right corner of (c)].

The DNA was mounted by the aqueous technique and stained with uranyl acetate.

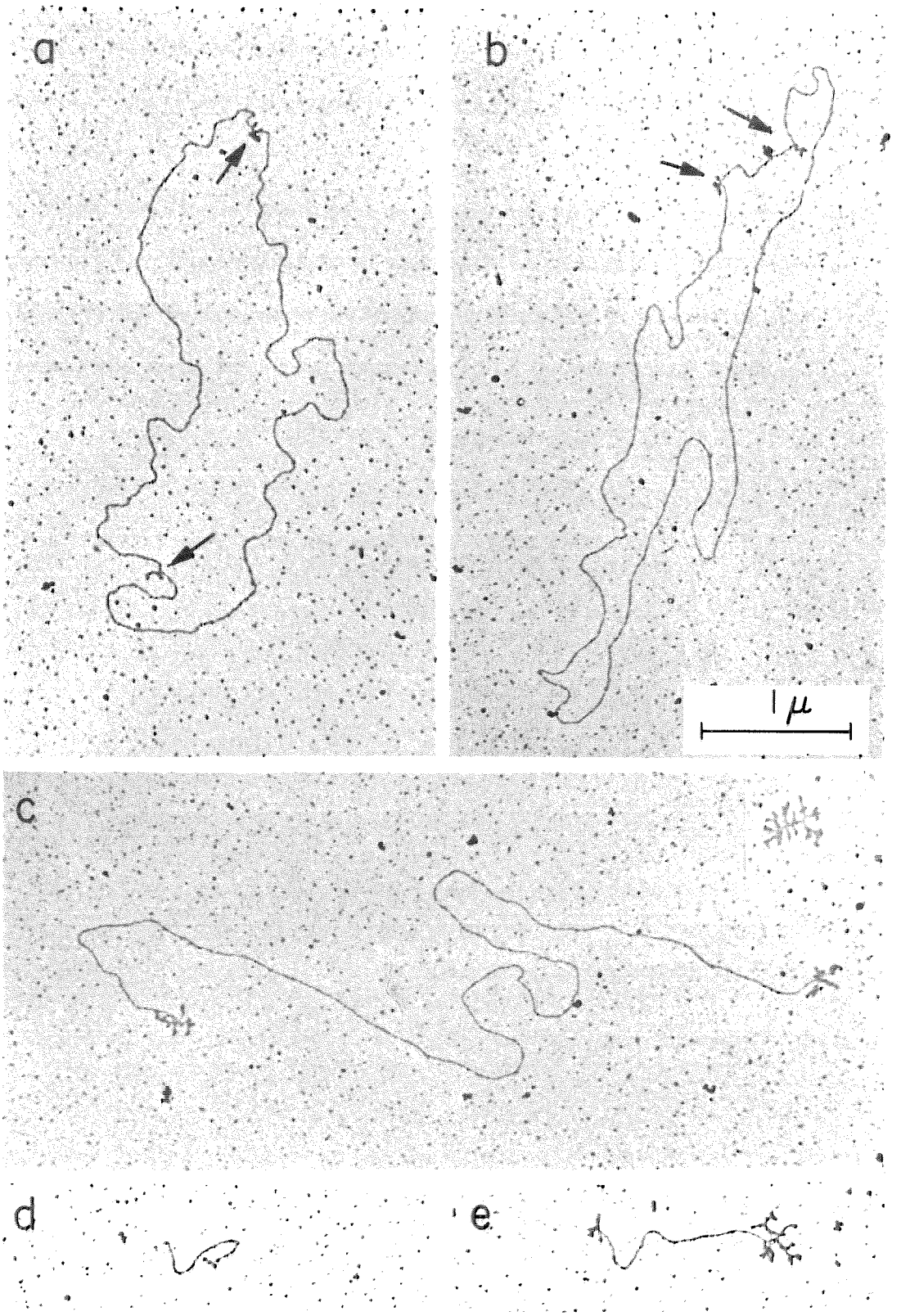

a
b
1 μ
c
d
e

Plate IV. Renatured coliphage 15 DNA with the single-stranded repetitious branches extended by formamide.

Two renatured circular coliphage 15 (WT) DNA molecules. The arrows mark the single-stranded repetitious branches. All whole circular molecules observed showed two such branches. The DNA was mounted by the formamide technique, stained with uranyl acetate, and shadowed with Pt-Pd as described in Materials and Methods.

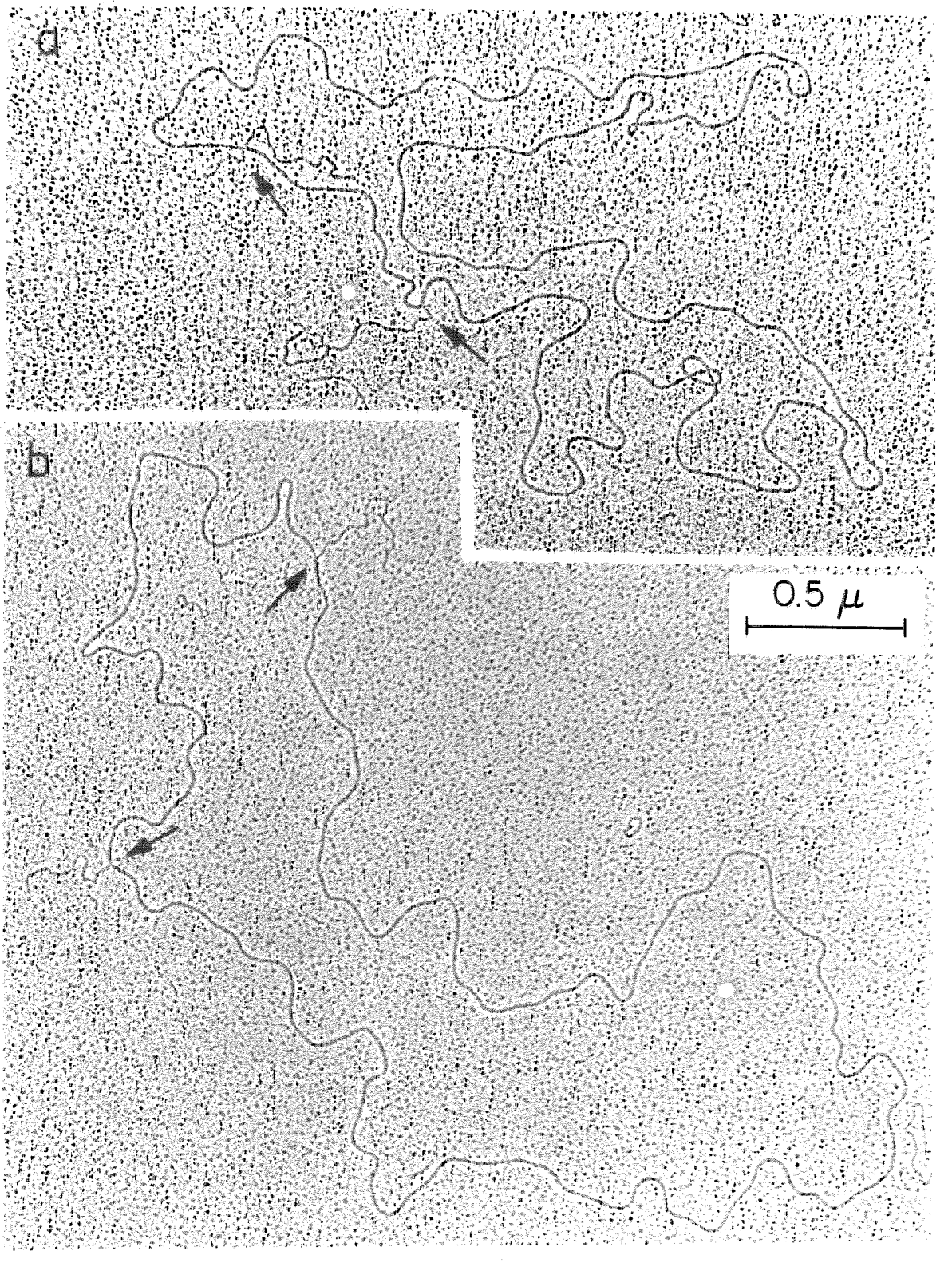

a
b
0.5 μ

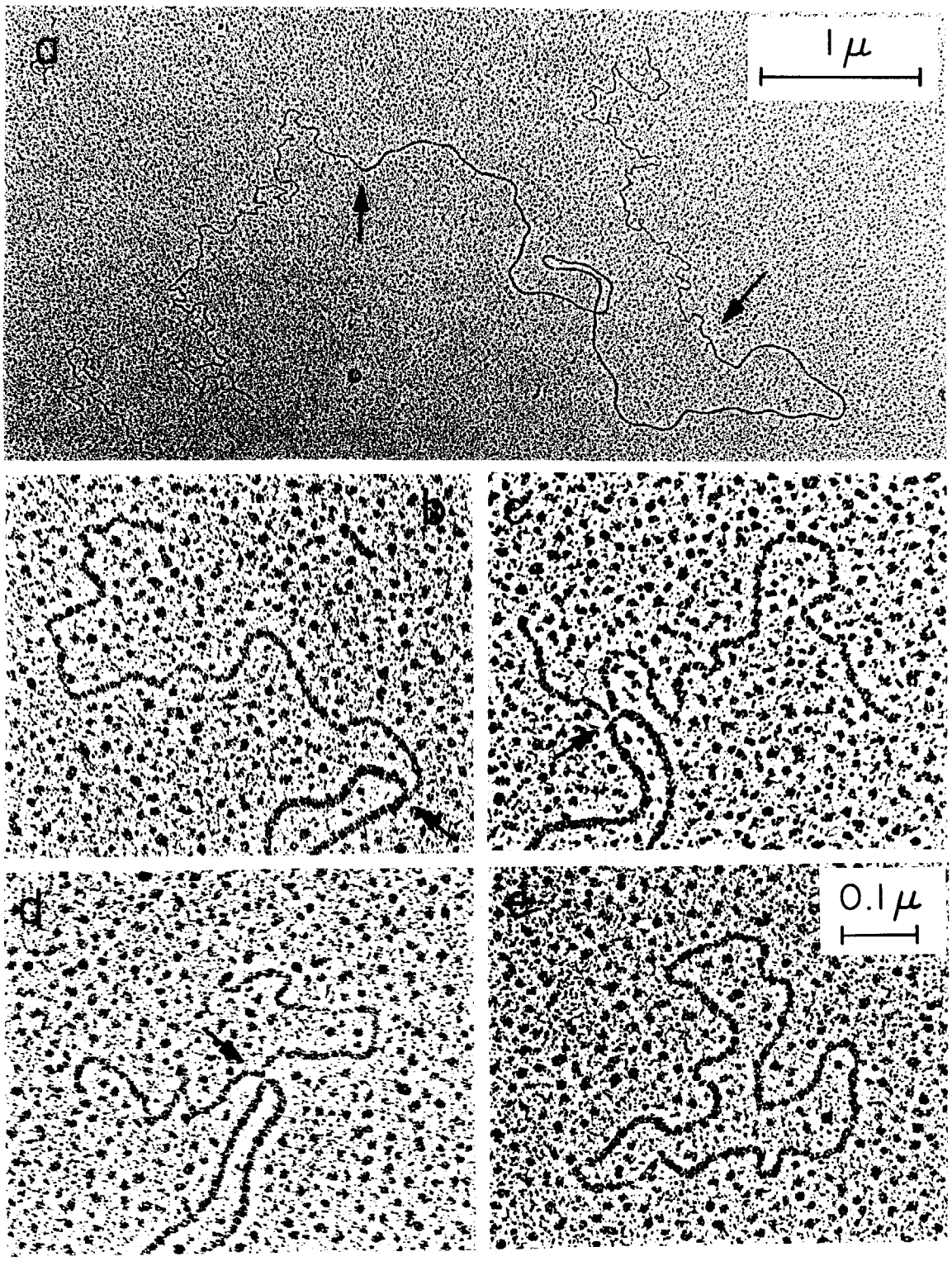
Plate V.

(a) Renatured linear coliphage 15 (WT) DNA molecule with single-stranded ends: The arrows mark the approximate junction points between single-stranded and double-stranded DNA.

(b), (c), and (d) Enlargements of unforked (b) and forked [(c) and (d)] branches.

(e) Single-stranded φX 174 DNA used for the calibration of the length measurements of the repetitious branches shown in (b), (c), and (d).

The φX DNA was mounted on the same grid that was used for Plates IV and V (b), (c), and (d).


a
1 μ
b
c
d
e
0.1 μ

solution so that the single-strand ends are extended into visible filaments which we call branches. A number of linear molecules with single-strand ends are also seen. [Plate III (c) and Plate V (a)]. Under our experimental conditions over 50% of the DNA renatured; the relative amounts of circular and linear species in the renatured preparations were about 40% and 60%, respectively.

If the mature phage molecules are circularly permuted, the contour lengths (circumferences) around the circles for the renatured molecules should be constant and equal to the contour length of the in vivo circular molecules. Histograms of the circumferences for the renatured molecules gave a mean length equal, within experimental error, to 11.9 μ, and a standard deviation of 0.4 μ.

We conclude that coliphage 15 DNA is indeed circularly permuted and terminally redundant.

(e) Distribution of beginning points for permuted sequences

We wish to investigate the distribution of beginning points in the ensemble of circularly permuted linear molecules from the mature phage. Information pertaining to this distribution can be obtained from measurements of the length of the double-strand regions in Form II molecules and the separation of bushes in the Form III molecules.

Let C be the length of the nonrepetitious sequences of the DNA molecules, and let T be the length of the terminal repetition. The length of the mature linear molecule is then C + T. Let p be the phase

difference between the beginning points of two strands which have reassociated, by which we mean that the physical distance between their beginning points is $pC$.

For Form II molecules the length, $D$, of the double-stranded region is then

$$D = C(1 - p) + T \qquad (1)$$

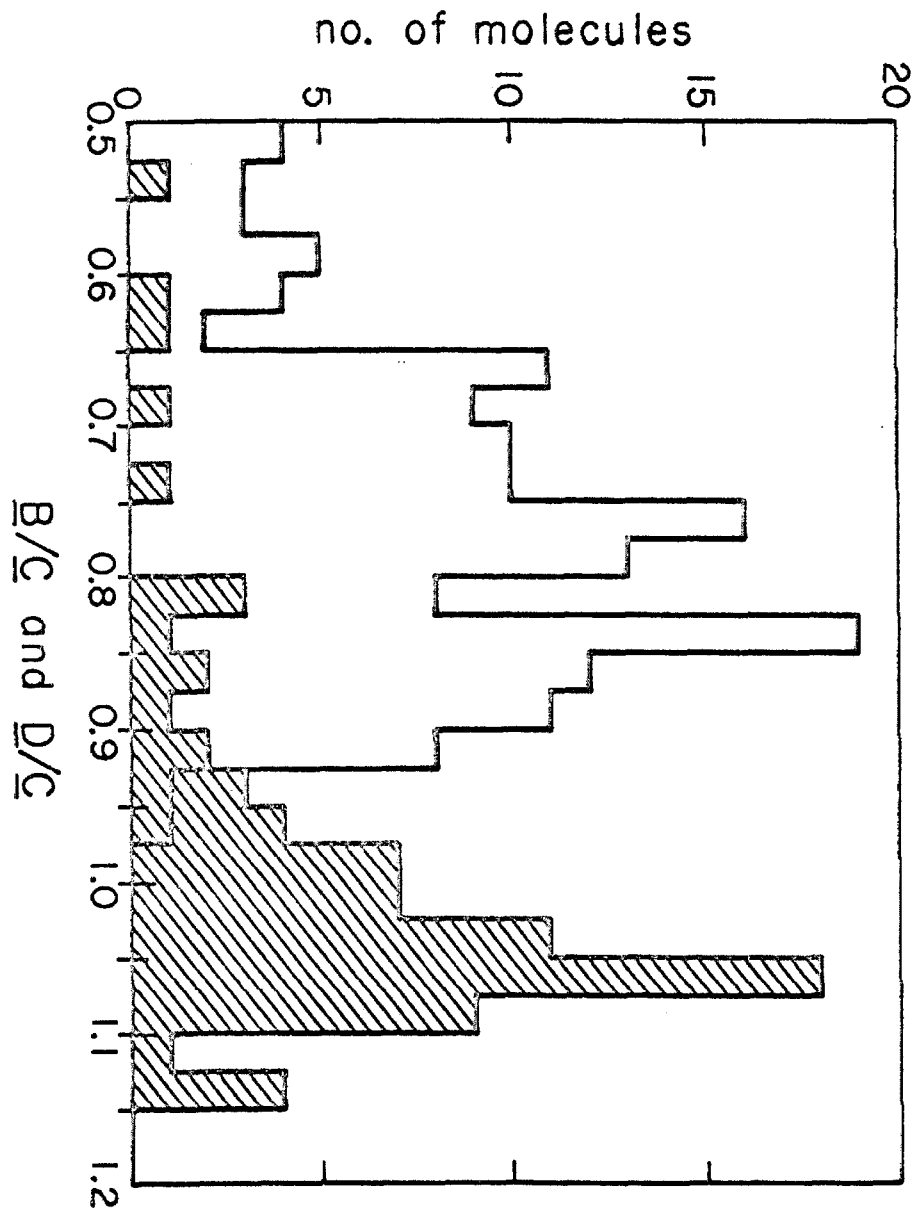
A Form II molecule has complementary ends of length $pC - T$ if $pC > T$. Such molecules are presumably, capable of cyclization.

In Form III molecules the circumference of the circle has a length $C$. The protruding single-strand branches have a length $T$. For a given $p < 0.5$ the separation of the branches in the shorter direction around the circle varies between $pC - T$ and $pC + T$ (depending on the phenomenon of branch migration as discussed later). Therefore, the separation of the branches in the longer direction around the circle, $B$, varies in the interval

$$C(1 - p) + T > B > C(1 - p) - T \qquad (2)$$

We use the longer distance around the circle as a measure of the branch separation because it is more directly comparable to the quantity $D$ of Form II molecules. The distribution of $p$ values may conceivably then be inferred from the observed distribution of $D$ and $B$ values. A histogram of observed $D/C$ and $B/C$ values is shown in Fig. 4. Among Form III (circular) molecules there are a few with $B/C$ values close to 0.5 but the number increases irregularly to about

Fig. 4. Distribution of the relative distance between single-strand branches on renatured molecules. All observations were made on molecules mounted by the aqueous technique so that the single-strand branches appeared as "bushes". A histogram (shaded area) of the length of the double-stranded portion between the end bushes of 78 linear molecules is given with the horizontal coordinate, $\underline{D}/\underline{C}$, being the length of the double-stranded portion between the end bushes of the linear molecules divided by the mean length of the renatured circular molecules. Similarly, a histogram of the quantity $\underline{B}/\underline{C}$ which is the longer distance around the circle between the two bushes of a renatured circular molecule divided by the circumference of that circular molecule is displayed (unshaded area). One hundred fifty circular molecules and 78 linear molecules from the photographs of coliphage 15 (TAU) were chosen at random for this plot. A similar plot was obtained for coliphage 15 (WT) DNA. To obtain a valid distribution for all molecules, the value of the vertical coordinate for the linear molecules in the plot above should be multiplied by 3 since they actually represent about 60% of the total population of renatured molecules, but only 78 out of 228 were scored for the histogram. The distributions are plotted in intervals of approximately 1 $\sigma$ relative to the observed values of $\underline{C}$.


no. of molecules
0
5
10
15
20
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
B/C and D/C

0.85 and then falls to a very low value for B/C close to 1.0. For linear molecules there are many with D/C values greater than 1.0, as expected if there is terminal repetition, and a decreasing number with decreasing D/C values.

The fact that there are some renatured molecules with B/C or D/C values in all intervals between 0.5 and 1.0 shows that a number of coordinates of beginning points, corresponding to p values between 0.0 and 0.5, occur in the population. A discussion of other features of the shape of the histogram in Fig. 4 will be given later.

(f) Size and heterogeneity of the terminal repetition

In previous physical work, the size of the terminal repetition of a DNA has been estimated mainly by the E. coli exonuclease III digestion method (MacHattie, et al., 1967; Ritchie, Thomas, MacHattie, and Wensink, 1967; Rhoades, et al., 1968). More accurate values can be derived by electron microscopy.

We estimate the size of the terminal repetition as 0.075(±0.005) C from the difference in lengths of the mature linear and the in vivo closed circular DNA's.

However, the most accurate values are obtained by direct measurements of single-strand branch lengths on micrographs of Form III molecules mounted under conditions in which the single-strand DNA is extended. For measurements of this type, it is essential that a single-strand DNA of known length be mounted on the same grid as an internal standard.

The phenomenon of branch migration should be mentioned before the branch length measurements are discussed. Electron micrographs of the branches due to terminal repetitions are shown in Plates IV and V. In some cases each branch consists of one single strand with one free end. In other cases the branch has forked into two sub-branches and there are two free ends. A schematic presentation of the mechanism of forking is shown in Fig. 5. On consideration of the physicochemical forces involved, it is expected that branch migration and forking will indeed occur, as observed.

The total branch length is the sum of the lengths of the two sub-branches. The relative fork position is defined as the length of the longer sub-branch divided by the total branch length. The probability of different fork positions has been studied by plotting the histogram of relative fork positions shown in Fig. 6. It appears that all fractional fork positions are equally probable, but unforked branches occur in about 20% of the cases and thus with greater than random probability.

The distribution of length of the branches and of the $\varphi$X 174 standards is plotted in Fig. 7. The ratio of the number average molecular length of the branches to the number average molecular length of $\varphi$X 174 is 0.536. (The corresponding ratio of weight average lengths is 0.554.) Since the length of the mature linear coliphage 15 DNA is 7.81 times that of double-stranded $\varphi$X 174 RF II DNA, the number average length of the terminal repetition is $0.536/(7.81 - 0.536) = 0.074$ of the length, $\underline{C}$, of the nonrepetitious part of the mature phage

Fig. 5. Schematic representation of the mechanism of branch migration. An enlargement of a portion of the circular Form III molecule in the lower right of Fig. 3 is shown. Since a protruding single-strand branch in the top figure has the same base sequence as an adjacent strand which is in the duplex, the branch can fork and migrate without changing the total number of base pairs.

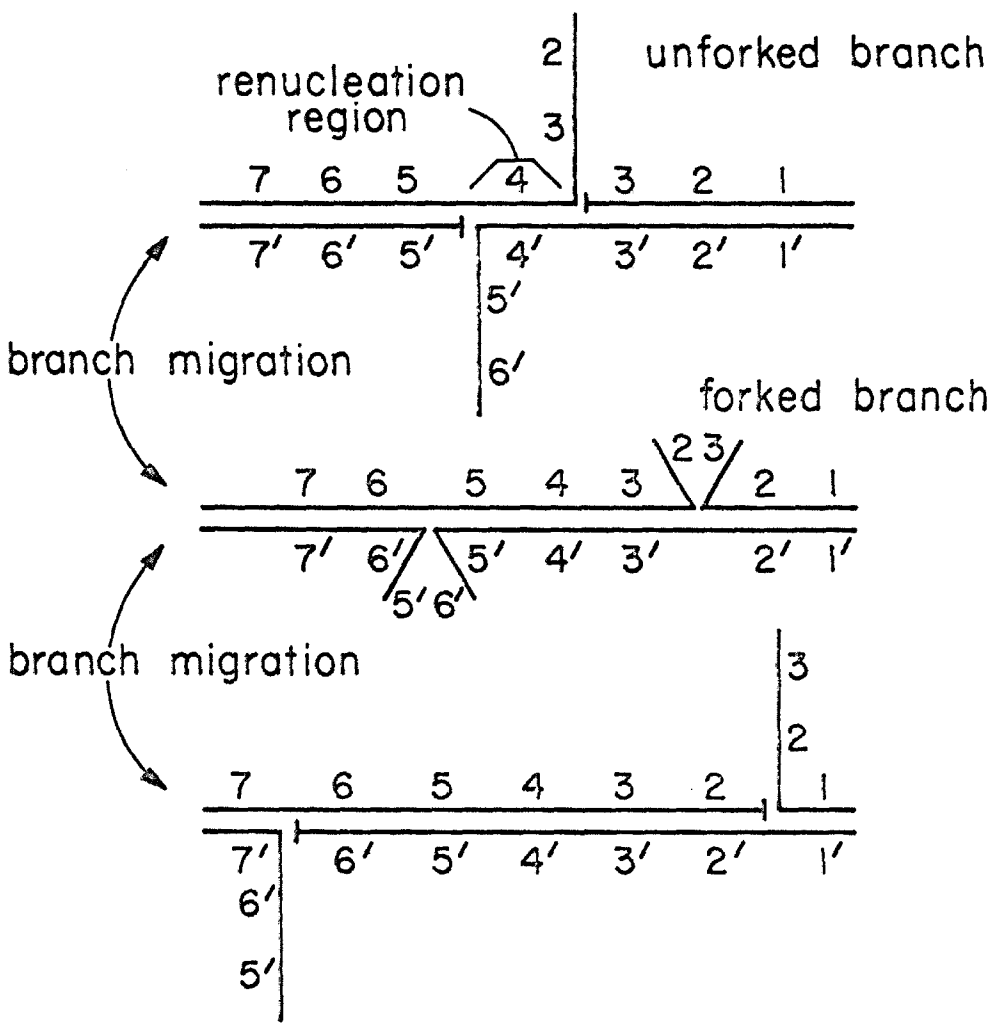

renucleation region
unforked branch
branch migration
forked branch
branch migration

Fig. 6. Relative fork position of repetitious branches. Coliphage 15 (WT) DNA was denatured and renatured for 6 hrs in 50% formamide at 25° C. The DNA was mounted for electron microscopy in the presence of formamide and the length of the repetitious branches measured. The relative fork position is evaluated by dividing the distance from the fork position to the end of the longest branch by the total length of the branch. Therefore, a relative fork position of 1.0 would represent a linear branch, while a relative fork position of 0.5 would represent a forked branch in which the lengths of the two single strands are equal. The distribution of the relative fork position for 91 repetitious branches is plotted here.

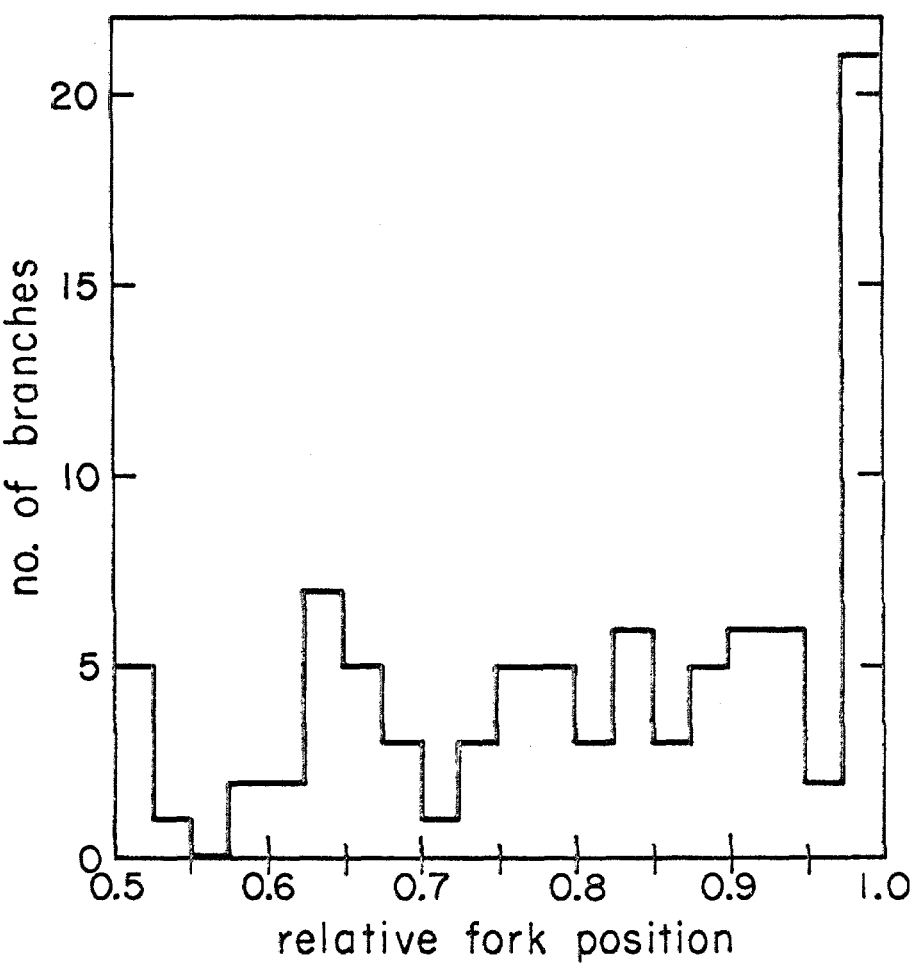
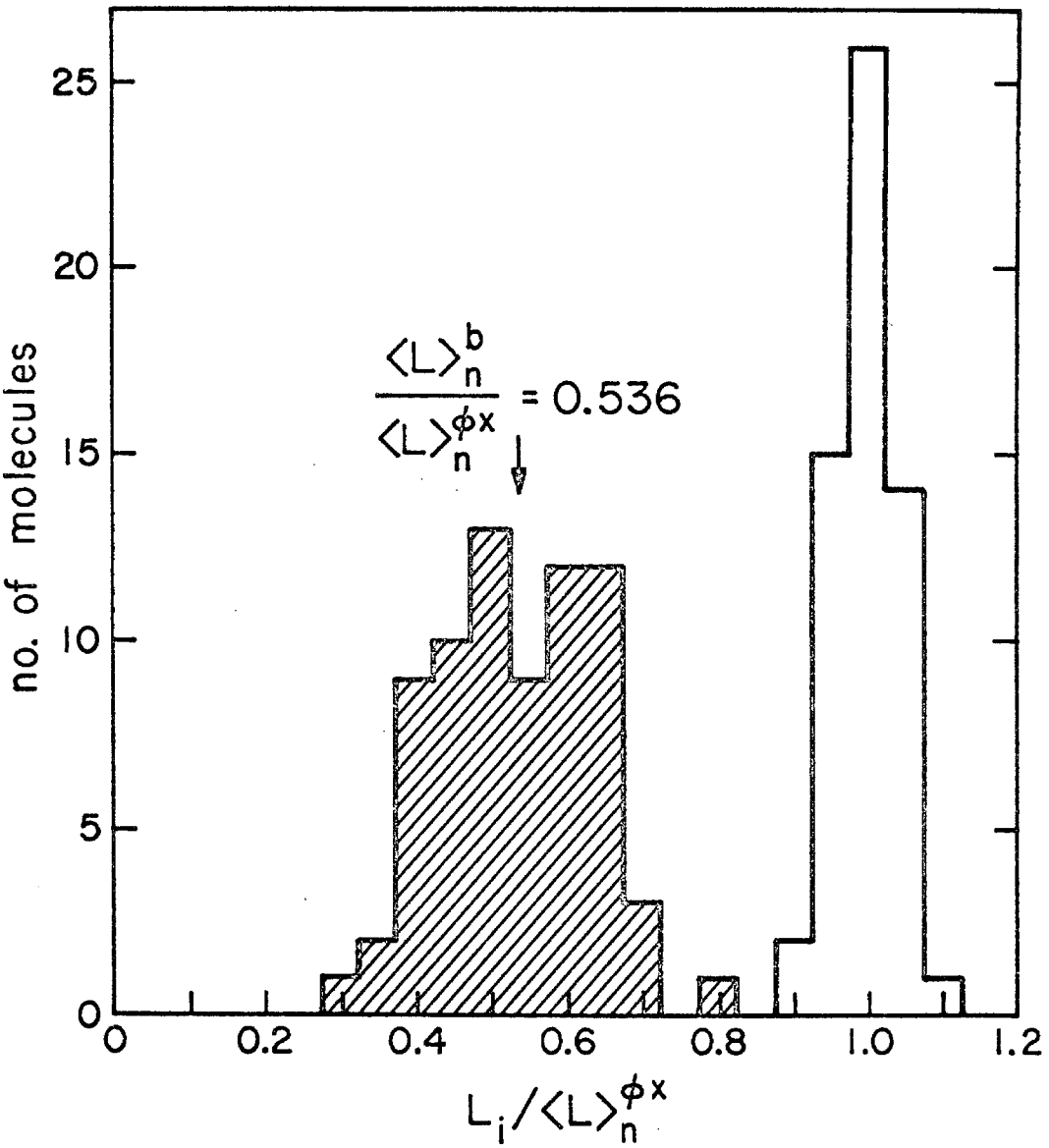

20
15
10
5
0
no. of branches
0.5
0.6
0.7
0.8
0.9
1.0
relative fork position

Fig. 7. Length distribution of the repetitious single-stranded DNA branches relative to φX 174 DNA. Renatured coliphage 15 (WT) was mounted along with single-stranded φX 174 DNA in the presence of formamide. Micrographs of both the repetitious single-stranded DNA branch (shaded area) and φX 174 DNA (unshaded area) were alternately taken. Only those renatured circular DNA molecules which showed two unambiguous repetitious single-stranded DNA branches are included in this distribution. The quantities $\underline{L}_i$, $\langle \underline{L} \rangle_n^{\varphi X}$, and $\langle L \rangle_{\underline{n}}^{\underline{b}}$ are contour length of a branch or molecule, number average length of φX, and number average length of the branches, respectively.


25
20
15
10
5
0
no. of molecules
$\frac{\langle L\rangle_n^b}{\langle L\rangle_n^{\phi x}} = 0.536$
0
0.2
0.4
0.6
0.8
1.0
1.2
$L_i/\langle L\rangle_n^{\phi x}$

DNA. This value agrees well with the less exact value deduced previously. It corresponds to 2800 base pairs.

It is obvious from Fig. 7 that the fractional standard deviation of the branch lengths is greater than the fractional standard deviation of φX 174 lengths. The fractional standard deviations are 0.19 for the branches and 0.043 for φX DNA. We attribute this large fractional standard deviation for the branch lengths to a true heterogeneity in length of the terminal repetition.

In order to discuss the heterogeneity quantitatively, we must first discuss the distribution in lengths for a homogeneous DNA population in electron micrographs. It is observed (Davis, et al., 1969) that for a homogeneous DNA population, the square of the standard deviation of a distribution of contour lengths on a micrograph is proportional to the mean length of the DNA. In fact, it is found that

$$(\sigma_d/\underline{L}_d) = 0.037(\pm 0.007)/\underline{L}_d^{1/2} \qquad (3)$$

and

$$(\sigma_s/\underline{L}_s) = 0.049(\pm 0.12)/\underline{L}_s^{1/2} \qquad (4)$$

where $\underline{L}$ represents contour length and subscripts $\underline{d}$ and $\underline{s}$ refer to double- and single-stranded DNA, respectively. The numerical constants in the equation above apply when double- and single-strand lengths are measured in units of the length of φX RF II double-stranded DNA and the length of φX single-stranded DNA, respectively. These relations were deduced from studies of DNA's with a large

range of lengths and apply under a wide range of mounting conditions. We believe that the primary sources of fluctuation in length occur in the structure of the DNA-protein film, and not in magnification or measuring errors.

The predicted fractional standard deviation for $\varphi$X 174 lengths from eq. 4 is 0.05(±0.01), in satisfactory agreement with the value observed for this sample of 0.043.

The predicted fractional standard deviation for a homogeneous population with the mean length of the terminal repetition is 0.07 (±0.01). Therefore we believe that the fractional standard deviation in length of the terminal repetition due to its heterogeneity is $(0.19^2 - 0.07^2)^{1/2} = 0.18$, or 500 base pairs. A variation in length of the terminal repetition of this magnitude corresponds to a fractional variation of 0.013 in length of the mature phage DNA, which is too small to be reliably observed.

(g) DNA-DNA homology studies by electron microscopy

The phage from wild type and TAU strains of E. coli 15 are morphologically similar, except that the latter have no tails. The DNA's have the same contour length and the same circular permutation-terminal repetition characteristics. We now ask the question as to the degree of sequence relationship between the DNA's.

A mixture of coliphage 15 (WT) and (TAU) DNA's was denatured, renatured, and mounted for electron microscopy in the presence of formamide. The resulting mixture will contain homoduplexes of the two types and, if there is any significant sequence homology, hetero-

duplexes with one strand each from the two kinds of DNA. In such heteroduplexes, regions of nonhomology are visible as single-strand loops of various kinds (Davis and Davidson, 1968; Westmoreland, et al., 1969; Davis, et al., 1969). In our experience, deletions, additions, or substitutions of nonhomologous regions as small as 50 to 100 base pairs are observable by this method. Point mutations cannot be detected.

No nonhomology regions were found by electron microscopic examination of the renatured mixture. Therefore, either there is no significant sequence relationship between the two DNA's so that no heteroduplexes form or the relationship is very close, within the limits stated above.

The hypothesis of no sequence relationship at all was eliminated by hybridizing a mixture of sheared wild type phage DNA and unsheared TAU DNA under conditions that only a relatively small amount of renaturation occurs. This method has been described previously in connection with a similar problem concerning several M. lysodeikticus phages (Lee and Davidson, 1969). If there is any sequence homology between the sheared and the unsheared DNA, heteroduplexes, each of which consists of a sheared piece of DNA and an unsheared strand, will form. These heteroduplexes can be recognized as molecules with a short inner double-strand region and two single-strand ends, one or both of which is large. The DNA is trapped in the cytochrome film from aqueous solution, so that single-strand regions appear as bushes.

The length of a single-strand piece can be approximately estimated from the perimeter of the bush (Davis and Davidson, 1968). Plate III (d) and (e) show relevant micrographs; the scoring of the observed grids is presented in Table 2. The molecules of interest are the whole/fragment heteroduplexes [Plate III (e)]. These are readily distinguished from whole/whole homoduplexes, which are either circular, or linear with short equal size end bushes and most frequently with a long double-strand region. Fragment/fragment homoduplexes have only short end bushes [Plate III (d)].

According to the steric hindrance effect in renaturation, which is explained in more detail in a later section, a fragment which is homologous to a region close to the topological center of a whole strand has a lower probability of penetrating the random coil and forming a heteroduplex than does a fragment which is homologous to a region close to the topological outside of the whole strand. The observations are in agreement with this theory and the whole/fragment heteroduplexes most frequently have one large and one small end bush. The occurrence of whole/fragment homoduplexes due to the presence of single-strand breaks in the unsheared TAU preparation can only be evaluated by the control sample of TAU alone.

It is clear from the scoring in Table 2 that whole/fragment heteroduplexes do occur in the large amount expected in, and only in, the renatured mixture of sheared wild type and unsheared TAU DNA. We therefore conclude that the two kinds of DNA are closely related and there are no sequence dissimilarities of length greater than 50 to 100 base pairs.

TABLE 2

Numbers of different kinds of molecules in heteroduplex preparation

| Classes of molecules | No of molecules | |
|---|---|---|
| | Whole TAU | Whole TAU + sheared WT |
| Long linear or cyclic double-stranded region with small equal sized end bushes (A) | 80 | 14 |
| Short double-stranded regions with large equal sized end bushes (B) | 2 | 3 |
| Long double-stranded region with small unequal sized end bushes (C) | 10 | 3 |
| Short double-stranded region with large unequal sized end bushes (D) | 9 | 66 |
| Short double-stranded region with small end bushes (E) | 11 | 192 |
| Whole single-stranded molecules (F) | 99 | 14 |
| TOTAL | 211 | 292 |
| Probable identification of scored molecules | | |
| Whole/whole homoduplexes = A + B | 82 | 17 |
| Whole/fragment homoduplexes = C | 10 | 3 |
| Whole/fragment heteroduplexes = D | 9 | 66 |
| Fragment/fragment homoduplex = E | 11 | 192 |
| Percent heteroduplexes* | 5% | 64% |

* The percentage of whole DNA strands hybridized to a small fragment = $100\ D/[2(A + B) + C + D]$

## 4. Recapitulation and Further Discussion

### (a) Phage properties

Auxotrophic mutants of wild type E. coli 15 have in general been derived by uv mutagenesis and the standard penicillin selection procedure. E. coli 15 $T^-$ was so derived (cited by Barner and Cohen, 1954); it was the mother strain for further uv mutagenesis and selection to produce E. coli 15 TAU (Barner and Cohen, 1958) and TAU-bar.* Of these four strains, only TAU-bar does not produce phage particles after uv or mitomycin C induction. [Our observations on this point agree with those of Cozzarelli, et al. (1968), and disagree with those of Bolton, et al. (1964).] Furthermore, the in vivo closed circular DNA species was found after induction in 15 WT and 15 TAU, but not in TAU-bar. It is possible then that either TAU-bar has been cured of the prophage or the prophage has mutated to a noninducible form by the uv mutagenesis.

We noted that in all strains studied there was a considerable decrease in the number of the 0.67 μ plasmid DNA circles after mitomycin C treatment.

Whereas the lysates of induced E. coli 15 WT and $T^-$ contained phage with attached tails and excess free tails, the TAU lysate con-

---

* We are informed by Professor P. C. Hanawalt that the TAU-bar was derived by R. Wax from an auxotrophic mutant E. coli 555-7 that R. Weatherwax originally isolated from 15 $T^-$. It was named TAU-bar by P. C. Hanawalt. It has also been named WWU by S. Person.

tained only tailless phage and no free tails. The studies reported here show that the DNA's are very similar in base sequence, and possibly differ only by some point mutations. Presumably, one or several such mutations have made the TAU DNA defective for the production of tails.

The coliphage 15 (WT) and ($T^-$) particles are colicinogenic to 15 WT and 15 $T^-$ bacteria; coliphage 15 (TAU) particles are not. These observations provide further evidence that tail particles are the colicinogenic component (Endo, et al., 1965).

Both TAU and TAU-bar bacteria are resistant to the killing action of the (WT) and ($T^-$) phage. These strains were isolated after uv irradiation of $T^-$ which induces the prophage; the selection procedure would therefore favor selection of a resistant strain.

In agreement with previous workers (Sandoval, et al., 1965) Mennigmann, 1965; Endo, et al., 1965) we have not succeeded in finding a host on which the phage particle produces a productive infection; the authors cited have suggested that the phage is defective. Our studies, however, show that the phage particles have a unique circularly permuted, terminally repetitious sequence with a complexity approximately equal to the molecular weight. Thus, this phage is not like the defective B. subtilis phage (Okamoto, Mudd, and Marmur, 1968) which contains a random sample of the bacterial chromosome as its DNA. Furthermore, Cowie and Szafranski (1965) have reported that

coliphage 15 (TAU) DNA has only 6% homology with host DNA.

(b) Properties of the circular permutation and terminal repetition

The DNA's of coliphage 15 (WT) and (TAU) are identical in size and base sequence by the criteria available.

Like the virulent phages T2 and T4 and the temperate phages P22 and P1 (Thomas and Rubinstein, 1964; Thomas and MacHattie, 1964; MacHattie, et al., 1967; Rhoades, et al., 1968; Thomas, et al., 1968; Ikeda and Tomizawa, 1968), the DNA of the E. coli 15 phages is terminally repetitious and circularly permuted.

The mean size of the terminal repetition is 0.074 of the whole genome, or 2800 base pairs. For T2 DNA, the size is reported to be 2000 to 5000 base pairs or 0.01 to 0.03 of the genome size; for P22 it is roughly 1000 base pairs or 0.025 of the genome size (Thomas, Kelly, and Rhoades, 1968); for P1, it is reported to be 0.12 of the genome or 10,000 base pairs (Ikeda and Tomizawa, 1968).

The accurate electron microscope methods used here have made it possible to observe that the size of the terminal repetition is somewhat variable, with a fractional standard deviation of 0.18 of the mean size of the terminal repetition, or 500 base pairs, or 0.013 of the genome size. This variability shows that the mechanism for cutting the DNA for packaging in the phage head is not precise. In particular, the variability is plausible for the "headful" mechanism of packaging DNA (Streisinger, Emrich, and Stahl, 1967). The cutting mechanism is, however, sufficiently precise so that every phage

has a full complement of genes.

After induction and before phage maturation, there is an intracellular closed circular DNA and the mature phage DNA is 0.075 longer than this intracellular form. The amount of this DNA increases from 15 min to 30 min after induction. It is highly probable that this closed circular molecule contains one complete genome of the coliphage 15 DNA with no terminal repetition.

### (c) Excluded volume effects and the renaturation of circularly permuted DNA's

In order to explain the observed dependence of the rate of renaturation on strand length, Wetmur and Davidson (1968) suggested, as one of several possibilities, that there is an excluded volume effect so that points near the physical center of a random coil are less available for the initial joining event in renaturation than are points close to the outside. Random walk theory shows that points near the topological center of a random coil have a smaller radius of gyration with respect to the center of mass of the coil than do points near the topological outside of the coil. Therefore, a point near the topological end of a strand is, on the average, closer to the physical outside of the coil and is more available for the initiation of a renaturation reaction than is a point close to the topological center of the strand.

This hypothesis has been confirmed by Wulff, Jamieson, and Davidson (1969) who showed that the rate of renaturation of sheared fragments of λDNA with whole strands was greater for sheared strands

homologous to points near the topological outside of the whole strand. Essentially the same observations by the same method are reported here in Table 2 in our study of the renaturation of sheared fragments of coliphage 15 (WT) DNA with the whole strands of coliphage 15 (TAU) DNA.

For two complementary whole strands with widely spaced beginning points (renaturing to duplexes with a relative p value close to 0.5 in eq. 1 and a relative branch separation close to 0.5 in Fig. 4) in a population of circularly permuted molecules, the outside of one molecule is the interior of the other, and vice versa; whereas, for two strands with beginning points close to each other (relative p values close to zero and a branch separation close to 1.0) the topological outsides are homologous. According to the excluded volume hypothesis, two strands of the latter type will renature more rapidly. We believe that the excluded volume phenomenon will have a significant effect on the relative rates of formation of molecules with different p values, but we do not have available an independently derived theoretical or experimental quantitative expression for the magnitude of the effect. In the discussion that follows, we shall assume that the probability of renaturation of two strands decreases linearly as a function of their relative p value, from a value of 1.0 for $p = 0$ to a value of 0 for $p = 0.5\,[1 + (T/C)]$ (this being the maximum possible value of p).

(d) The physical chemistry of branch migration

Consider a molecule with a forked branch. The exact position of the fork does not affect the total number of base pairs in the duplex or

the steric interaction at the point of the fork. Thus, it is reasonable that all fork positions are equally probable. There are 2800 ways to have a forked branch and two ways to have an unforked branch. Experimentally, the ratio of unforked to forked branches is 1:4. If this is an equilibrium ratio at 300° K, and in view of the statistical factor of 1:1400, there is an intrinsic difference in free energy between an unforked and a forked branch of 3500 cal/mole in favor of the unforked branch. The steric interactions at the junction of the two sub-branches are different than the steric interactions with only a single branch, and a free energy difference of the magnitude calculated is not unreasonable.

Let $\tau$ be the jump time for the migration of a branch by one nucleotide. The mean time to equilibrate the fork position over 2800 nucleotides is then of the order of $(2800)^2 \tau$. The sample scored in the histogram of fork positions (Fig. 6) was allowed to renature for about $10^4$ sec; thus an upper limit on $\tau$ is $10^{-3}$ sec, if equilibrium has indeed been achieved. There are no strictly comparable measurements in the literature for comparison, but this relaxation time is of the same order as that observed by Massie and Zimm (1969) for the relaxation time per nucleotide in ordinary renaturation under comparable conditions. Better measurements of the equilibrium and kinetics in branch migration would appear to be both useful and practicable.

(e) The distribution of beginning points

The question at issue is whether the circular permutation around the genome is complete, or whether there are certain points or regions

along the genome which have a higher than random probability of being beginning points.

The histogram of Fig. 4 does not show a uniform occurrence of molecules with all possible branch positions and hence, at first thought, suggests a nonuniform distribution of beginning points in the population of DNA molecules. We shall argue below that there are several physical factors which contribute to the shape of the histogram and that there is no evidence in Fig. 4 against the hypothesis of a uniform distribution of beginning points. With reference to Fig. 4, for renatured linear molecules with a double-strand length of $\underline{D}$, the separation $\underline{p}$ of the beginning points of the two strands in units of the genome size, $\underline{C}$, is $\underline{p} = 1 - (\underline{D}/\underline{C}) + \underline{t}$, where $\underline{t} = \underline{T}/\underline{C}$ is the fractional length of the terminal repetition. In Fig. 4 there are renatured linear molecules with values of $\underline{D}/\underline{C}$ in the interval 1.0 to $1.0 + \underline{t}$. These are molecules with $\underline{p}$ values in the interval $0 < \underline{p} < \underline{t}$. They do not have mutually complementary regions at their ends and are incapable of cyclization. The important point is that these are a class of molecules with beginning points of the two strands less than one terminal repetition apart.

There are some renatured linear molecules with values of $\underline{p} > \underline{t}$. These molecules presumably do have complementary regions on their ends and are capable of cyclizing but have not done so for kinetic reasons. The rate of cyclization is presumably proportional to the length of the complementary ends and therefore slower the shorter the ends. However, it should be noted that for all but an insignificant number of such molecules, the complementary ends are greater than 200 base pairs long and, according to arguments given previously (Wang and Davidson, 1966a), the equilibrium under renaturing conditions strongly favors cyclization.

For renatured circular molecules, we define the fractional branch separation by the shorter distance around the circle as $b = 1 - (B/C)$, where $B$ was defined previously as the separation of the branches by the longer way around the circle. In principle, molecules with values of $p > t$ are capable of cyclization. Due to branch migration, the branch separation, $b$, for a given value of $p$ can be anywhere between $p - t$ and $p + t$. As we shall see, this affects the distribution function for $b$ values.

The histogram of Fig. 4 shows an increase in the number of renatured circular molecules with decreasing $b$ values between 0.5 and 0.15 but a sharp fall in the number of such molecules as $b$ goes from 0.15 to 0. Thus, the results show that a number of positions along the chromosome occur as beginning points.

We now attempt to interpret the shape of the distribution curve of Fig. 4. The increase in the number of molecules with decreasing $b$ values between 0.5 and 0.15 can be attributed, partly or solely, to the selective renaturation phenomenon proposed in a preceding section. Furthermore, the sharp decrease in the number of molecules as $b$ goes from 0.15 to 0 can be attributed to branch migration, as discussed below. Thus, there is no evidence in the histogram against the hypothesis of a random distribution of beginning points.

The argument about the effect of branch migration goes as follows: Let $f(p)dp$ be the fractional number of circular molecules and molecules which are in principle capable of cyclization with $p$ values

in the interval $p$ to $p + dp$. The range for $p$ is $0 < p < 0.5(1 + t)$. (A renatured molecule with a $p$ value, $p_1$, greater than $0.5(1 + t)$ gives the same branch separation as though its $p$ value were $(1 + t - p_1)$ which is less than $0.5(1 + t)$. In fact, the experimental method used here cannot distinguish between circular permutations over half the chromosome and over all of the chromosome.) For a given $p$, the branch separation $b$ lies between $p - t$ and $p + t$ due to branch migration. We assume that complete equilibration of all branch positions has occurred and that for a given $p$ there is a uniform probability of any value of $b$ between $p - t$ and $p + t$. (The experimental fact is that there is a slightly higher probability of the values $p - t$ and/or $p + t$ than any of the intermediate values, but we neglect this.)

Let $N(b)^{db}$ be the number of renatured circular molecules with a branch separation in the interval $b$ to $b + db$. We wish to evaluate $N(b)$, given $f(p)$ and the phenomenon of branch migration.

A molecule with a branch separation $b$ such that $2t < b < 0.5\text{-}1.5\ t$ receives contributions from all $p$ values in the range $b - t$ to $b + t$. However, molecules with small $b$ values ($0 < b < 2t$) receive contributions only from a smaller interval of $p$ values between $t$ and $b + t$. This is the important effect which decreases the number of molecules with small $b$ values. The case for $b$ values greater than $0.5\text{-}1.5\ t$ is a little more complicated and not important here; the answer is correctly stated in the range of conditions given for the integrations below.

Assuming complete equilibration of all fork positions, the function $\underline{N}(\underline{b})$ is given by

$$\underline{N}(\underline{b}) = \int_{\underline{t}}^{\underline{b}+\underline{t}} \underline{f}(\underline{p})d\underline{p} \quad \text{for} \quad 0 < \underline{b} < 2\underline{t} \tag{5}$$

$$\underline{N}(\underline{b}) = \int_{\underline{b}-\underline{t}}^{\underline{b}+\underline{t}} \underline{f}(\underline{p})d\underline{p} \quad \text{for} \quad 2\underline{t} < \underline{b} < 0.5-1.5\underline{t} \tag{6}$$

$$\underline{N}(\underline{b}) = \int_{\underline{b}-\underline{t}}^{\underline{b}+\underline{t}} \underline{f}(\underline{p})d\underline{p} + \int_{1-\underline{b}-\underline{t}}^{0.5+0.5\underline{t}} \underline{f}(\underline{p})d\underline{p}$$

$$\text{for} \quad 0.5 - 1.5\underline{t} < \underline{b} < 0.5-0.5\underline{t} \tag{7}$$

$$\underline{N}(\underline{b}) = \int_{\underline{b}-\underline{t}}^{0.5+0.5\underline{t}} \underline{f}(\underline{p})d\underline{p} + \int_{1-\underline{b}-\underline{t}}^{0.5+0.5\underline{t}} \underline{f}(\underline{p})d\underline{p}$$

$$\text{for} \quad 0.5 - 0.5\underline{t} < \underline{b} < 0.5 \tag{8}$$

To calculate $\underline{N}(\underline{b})$ we need to know $\underline{f}(\underline{p})$. The function $\underline{f}(\underline{p})$ is affected by the actual distribution of beginning points and by the selective renaturation phenomenon. The latter causes $\underline{f}(\underline{p})$ to decrease with increasing $\underline{p}$. To calculate a theoretical curve, we assume that $\underline{f}(\underline{p})$ decreases linearly from $\underline{f}(\underline{p}) = 1$ at $\underline{p} = 0$ to $\underline{f}(\underline{p}) = 0$ for $\underline{p} = 0.5(1 + \underline{t})$ due to the selective renaturation phenomenon, that is,

$$\underline{f}(\underline{p}) = 1 - \underline{p}/0.5(1 + \underline{t}) \tag{9}$$

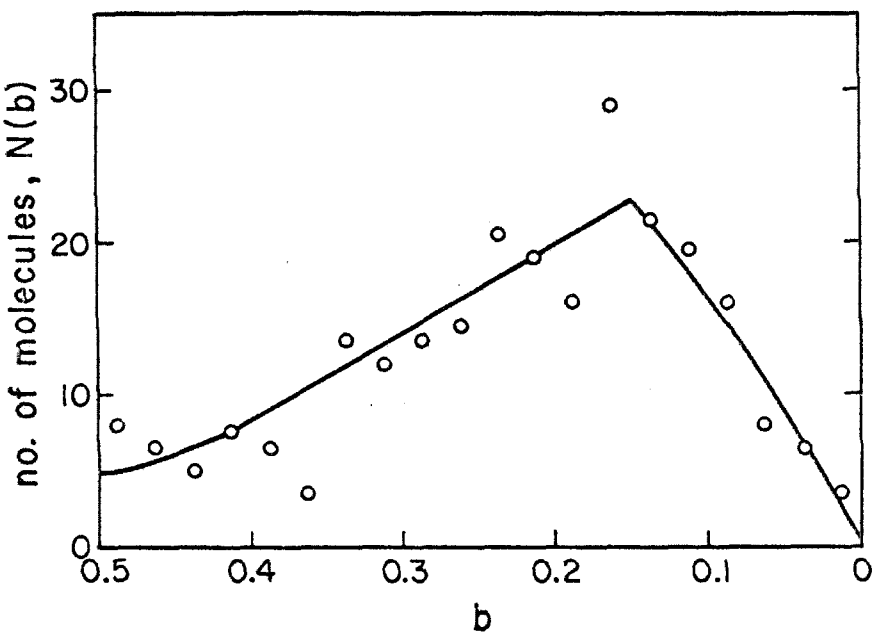
The curve so calculated from equations 5, 6, 7, and 8 is presented as a theoretical curve in Fig. 8. To compare it with the observed

histogram of Fig. 4, we must add a correction to the latter for the linear molecules with values of $p > t$ which have failed to cyclize. This correction is made by assuming that such a linear molecule with a given value of $p$ contributes uniformly to all values of $b$ between $p - t$ and $p + t$. The histogram of $b$ values from Fig. 4, so corrected, is shown as a set of points in Fig. 8. There is good agreement between the "theoretical" curve and the experimental points. The most important features of the curve, the maximum around $b = 2t$, and the decrease of $N(b)$ to zero as $b$ goes to 0, are not strongly dependent on the exact shape of the $f(p)$ curve, provided it is relatively uniform for $0 < p < 4t$.

In view of all of the experimental errors and the theoretical assumptions, we can only conclude that the present data are consistent with the idea of a uniform distribution of beginning points, with the observed distribution of branch separations modified by the selective renaturation phenomenon and branch migration.

Finally, we note that Ikeda and Tomizawa (1968) have reported a histogram which is qualitatively similar to that of Fig. 4 for the distribution of branch separations in renatured P1 DNA molecules. They observe an increase in the number of renatured molecules with decreasing $p$ values for $p > 0.1$ (approx.). Their tentative explanation that linear polymerization is more probable than cyclization for renatured Form II molecules with widely spaced beginning points and therefore with large complementary ends is in our opinion not acceptable, since the relative probability of cyclization and linear

Fig. 8. Comparison of "theoretical" (see text) and experimental distribution curve of branch separations, $\underline{b}$. The theoretical curve has been normalized to have about the same area as the experimental curve.


no. of molecules, N(b)
30
20
10
0
0.5
0.4
0.3
0.2
0.1
0
b

polymerization depends on the concentration of DNA and not on the length of the complementary ends which are equally available for both reactions (Wang and Davidson, 1966b).

The histogram presented by Ikeda and Tomizawa also shows a paucity of renatured cyclic molecules with branch separation less than the size of the terminal redundancy. Each of the several factors discovered above for coliphage 15 DNA may also be operative for P1 DNA.

Finally, we note that the phenomenon of branch migration and the experimental variability would make it difficult, by the methods used here, to recognize a distribution curve for beginning points with several maxima spaced one terminal repetition apart. Such a distribution is expected for models which involve replication of a moderately long concatenate starting at a specific origin of replication with the concatenate being cut in accordance with the headful mechanism at approximately constant lengths of one genome plus one terminal repetition.

Acknowledgements

We are grateful to Professor P. C. Hanawalt, Professor S. S. Cohen, and Dr. F. Funk for gifts of E. coli strains and related advice and information. We are grateful to Professor R. S. Edgar and Mr. J. S. Parkinson for advice and consultation and to Mrs. Chris Jamieson for her painstaking measurements of the electron micrographs. This research has been supported by Grant GM 10991 from the United States Public Health Service. One of us (RWD) has been supported by a predoctoral traineeship under Grant GM 01262. This work is contribution no. 3925 of the Department of Chemistry of the California Institute of Technology.

Barner, H. D. & Cohen, S. S. (1954). J. Bact. 68, 80.

Barner, H. D. & Cohen, S. S. (1958). Biochim. Biophys. Acta. 30, 12.

Bolton, E. T., Britten, R. J., Cowie, D. B., Roberts, R. B., Czafranski, P. & Waring, M. J. (1964). Carnegie Inst. Washington. Year Book 64, pp. 333-341.

Burgi, E. & Hershey, A. D. (1963). Biophys. J. 3, 309.

Cowie, D. B. & Szafranski, P. (1965). Carnegie Institution of Washington. Year Book 65, pp. 106-123.

Cozzarelli, N. R., Kelly, R. B. & Kornberg, A. (1968). Proc. Nat. Acad. Sci., Wash. 60, 992.

Davis, R. W. & Davidson, N. (1968). Proc. Nat. Acad. Sci., Wash. 60, 243.

Davis, R. W. (1969). Ph.D. Thesis, California Institute of Technology.

Davis, R., Simon, M. N. & Davidson, N. (1969). Manuscript in preparation for Methods in Enzymology.

Endo, H., Ayabe, K., Amako, K. & Takeya, K. (1965). Virology, 25, 469.

Frampton, E. W. & Mandel, M. (1968). Bact. Proc. 1, 153.

Ikeda, H. & Tomizawa, J-I. (1968). Cold Spr. Harb. Symp. Quant. Biol. 33, 791.

Kleinschmidt, A. K. & Zahn, R. K. (1959). Z. Naturforsch. 14b, 770.

Lee, C. S., Davidson, N. & Scaletti, J. V. (1968). Biochem. Biophys. Res. Commun. 32, 752.

Lee, C. S. & Davidson, N. (1969). Submitted to Virology.

MacHattie, L. A. & Thomas, C. A., Jr. (1964). Science 144, 1142.

MacHattie, L. A., Ritchie, D. A., Thomas, C. A., Jr. & Richardson, C. C. (1967). J. Mol. Biol. 23, 355.

Massie, H. R. & Zimm, B. H. (1969). Biopolymers 7, 475.

Mennigmann, H. D. (1965). J. gen. Microbiol. 41, 151.

Mukai, F. H. (1960). J. gen. Microbiol. 23, 539.

Okamoto, K., Mudd, J. A., Mangan, J., Huang, W. H., Subbaiah, T. V. & Marmur, J. (1968). J. Mol. Biol. 34, 413.

Okamoto, K., Mudd, J. A. & Marmur, J. (1968). J. Mol. Biol. 34, 429.

Parkinson, J. S. (1969). Personal communication.

Radloff, R., Bauer, W. & Vinograd, J. (1967). Proc. Nat. Acad. Sci., Wash. 57, 1514.

Rhoades, M., MacHattie, L. A. & Thomas, C. A., Jr. (1968). J. Mol. Biol. 37, 21.

Richardson, C. C. & Weiss, B. (1966). J. Gen. Physiol. 49, 81.

Ritchie, D. A., Thomas C. A., Jr., MacHattie, L. A. & Wensink, P. C. (1967). J. Mol. Biol. 23, 365.

Ryan, F. J., Fried, P. & Mukai, F. H. (1955). Biochim. Biophys. Acta. 18, 131.

Sandoval, H. K., Reilly, H. C. & Tandler, B. (1965). Nature 205, 522.

Sinsheimer, R. L. (1959). J. Mol. Biol. 1, 43.

Streisinger, G., Emrich, J. & Stahl, M. M. (1967). Proc. Nat. Acad. Sci., Wash. 57, 292.

Studier, F. W. (1965). J. Mol. Biol. 11, 373.

Thomas, C. A., Jr. & Rubenstein, I. (1964). Biophys. J. 4, 93.

Thomas, C. A., Jr. & MacHattie, L.A. (1964). Proc. Nat. Acad. Sci., Wash. 52, 1297.

Thomas, C. A., Jr., Kelly, T. J., Jr. & Rhoades, M. (1968). Cold. Spr. Harb. Symp. of Quant. Biol. 33, 417.

Wang, J. C. & Davidson, N. (1966a). J. Mol. Biol. 15, 111.

Wang, J. C. & Davidson, N. (1966b) J. Mol. Biol. 19, 469.

Westmoreland, B. C., Szybalski, W. & Ris, H. (1969). Science 163, 1343.

Wetmur, J. G. & Davidson, N. (1968). J. Mol. Biol. 31, 349.

Wulff, K., Jamieson, J. B. & Davidson, N., Unpublished results.

# CHAPTER 6

## Homology and Structural Relationships Between the Dimeric and Monomeric Circular Forms of Mitochondrial DNA from Human Leukemic Leukocytes

David A. Clayton, Ronald W. Davis and Jerome Vinograd

The following paper is in press in the Journal of Molecular Biology. My contribution to this work was the electron microscopy.

## 1. Introduction

One species of complex mitochondrial DNA, the circular dimer, occurs in human leukemic leukocytes (Clayton & Vinograd, 1967, 1969) and in human tumors (C. A. Smith, private communication), but does not occur in a variety of normal tissues (Clayton, Smith, Jordan, Teplitz & Vinograd, 1968). As the name implies, the circular dimer was considered to be a structure containing two mitochondrial genomes. This provisional assignment was based on the observations that the contour length of the dimer was twice the length of the monomer and that the base compositions of the monomer and dimer were similar. The latter was inferred from the observation that a mixture of unequal amounts of the two species formed a symmetrical band in a buoyant cesium chloride density gradient. In this communication we report the results of experiments designed to answer the questions: Does the dimer consist essentially of two mitochondrial DNA (M DNA) genomes linked in tandem and arranged in the form of a closed circular duplex? If so, does the dimer contain deletions, insertions, or regions of nonhomology?

In these studies we compare monomers and circular dimers that occur in mitochondrial DNA preparations from peripheral leukocytes of a patient with chronic myelogenous leukemia. The mitochondrial DNA from this patient, M. C., contained approximately 30% circular dimers, 60% monomers, and 10% catenated oligomers. We find that the buoyant densities of the separated monomers and circular dimers are identical within experimental error, a result

that indicates that the base compositions of the two duplexes are the same within ± 1.0%. The buoyant densities in alkaline cesium chloride of the corresponding heavy and light complementary strands were also the same within the experimental error. This result indicates that the base compositions of the corresponding separate strands in the two forms are likely to be very similar.

The degree of homology between monomer and dimer M DNA was examined by both buoyant density and electron microscope procedures. A high molecular weight hybrid was formed upon annealing approximately equal amounts of light monomer and heavy dimer strands. The buoyant density of the hybrid is indistinguishable from that of the native duplex, indicating that a degree of homology greater than 90% exists between the circular dimer and the monomer forms.

A mixture of singly nicked monomers and singly nicked dimers was successively denatured and reannealed in formamide solutions. The products were then examined in the electron microscope by the procedures developed by Davis & Davidson (1968), Davis, Simon & Davidson (1969), and Westmoreland, Szybalski & Ris (1969) to detect single-stranded loops and bushes. The electron micrographs showed that continuous regions of heterology, or deletions or insertions, greater than approximately 100 base pairs, were absent. A new kind of circular DNA structure resembling a "figure 8" was formed in these experiments. We call this DNA species which contains one dimer strand and two monomer strands a <u>fused dimer</u>.

A circular dimer can, in principle, be formed from two

monomeric genomes in two different ways. One method leads to a head-to-tail tandem arrangement of monomer units (Fig. 1(a)). The other leads to a head-to-head arrangement (Fig. 1(b)). We have used electron microscope procedures to distinguish between these models.

## 2. Materials and Methods

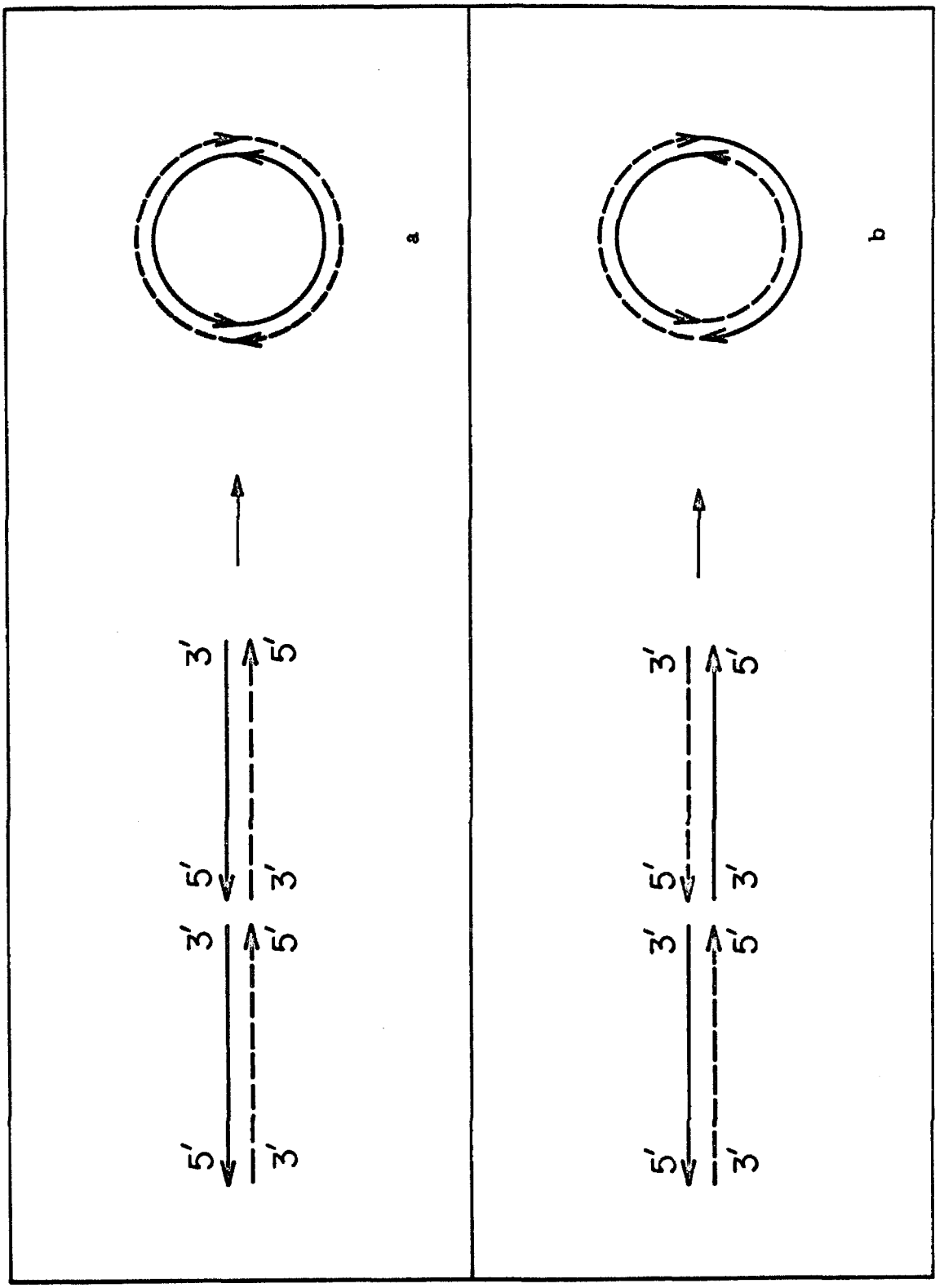
### (a) Isolation and purification of M DNA

Mitochondrial DNA was isolated from purified mitochondria of peripheral leukocytes obtained from a donor with untreated chronic myelogenous leukemia by methods reported previously (Clayton & Vinograd, 1969). The portions of the cesium chloride-ethidium bromide (CsCl-EB) gradients subtended by the upper and lower bands were collected at weekly intervals and pooled (fraction A). An alkaline buoyant analysis of fraction A showed that there was less than 5% nuclear DNA present in the pooled unfractionated M DNA sample. The pooled samples were again centrifuged to equilibrium and collected into an upper band (fraction B) and a middle and lower band (fraction C).

### (b) Separation of circular dimer and monomer M DNA

Circular dimers were separated from monomers by preparative sedimentation velocity in CsCl-EB (Watson, Bauer & Vinograd, 1969). Fraction B, containing about 3 $\mu$g of nicked M DNA, was

Figure 1. Representation of the two modes of formation of the circular dimer from two monomers. (a) Head-to-tail tandem arrangement of monomer units. (b) Head-to-head arrangement of monomer units.

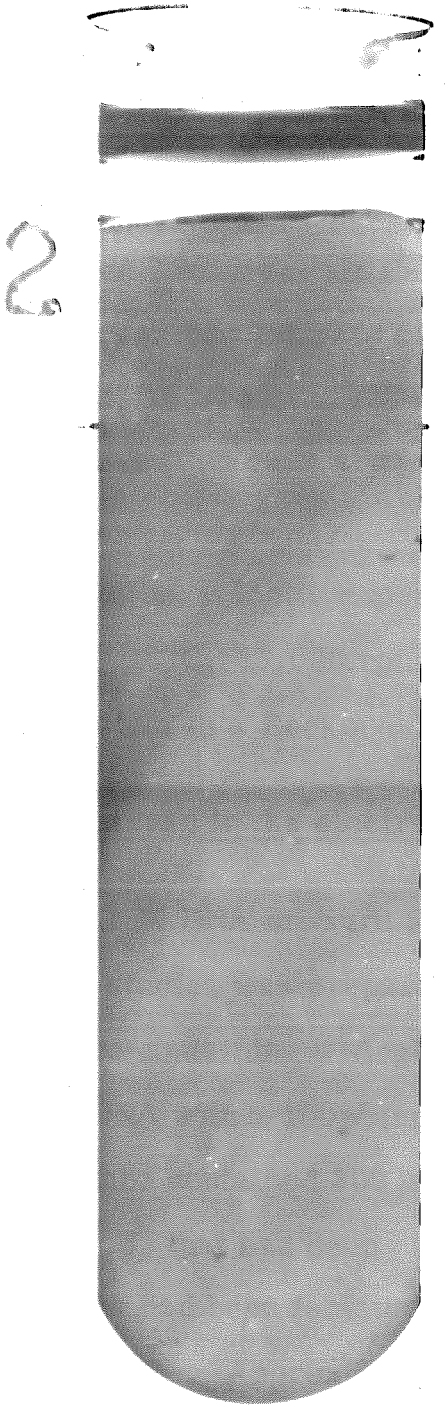

a
3'
5'
5'
3'
3'
5'
5'
3'
b
3'
5'
5'
3'
3'
5'
5'
3'

diluted to 5 ml. with 0.01 M-Tris, pH 7.5, and 0.001 M-EDTA, and centrifuged for 12 hr at 40 Krpm in an SW50.1 rotor. The M DNA pellet was allowed to resuspend in 0.4 ml. of supernatant and placed on 4.5 ml. of CsCl, $\rho$ = 1.40 g/ml., 0.01 Tris, pH 7.5, 150 $\mu$g/ml. EB, and centrifuged for $3\frac{1}{4}$ hr at 45 Krpm in an SW50.1 rotor. The 33 S nicked circular dimers and 26 S nicked monomers were detected by fluorescence and were separated by approximately 4 mm in the centrifuge tube (Fig. 2). The two M DNA species were isolated by drop collection and the EB removed by passage through a small column of Dowex-50 cation exchange resin. The small amount of heterogeneous nuclear DNA was removed in this preparative velocity experiment (Fig. 2), as is shown in the buoyant analyses of fractions $B_1$ and $B_2$ (Fig. 3).

Purified closed circular dimer DNA was obtained from fraction C in a similar sedimentation velocity experiment. The closed dimer (fraction $C_1$) was freed of dye and concentrated as described above.

Examination of the purified nicked circular dimers (fraction $B_2$) in the electron microscope revealed 85% circular dimers, 12% monomers, and 3% ambiguous and catenated dimers. A similar analysis of the purified nicked monomer sample (fraction $B_1$) gave 97% monomers and 3% circular dimers. These samples were used for the determination of the neutral and alkaline buoyant densities and for the hybridization experiments analyzed by centrifugation as described below.

Figure 2. The separation of nicked circular dimer (lower band) and nicked monomer (upper band) M DNA after velocity sedimentation in a cesium chloride-ethidium bromide solution. Centrifugation conditions are described in Materials and Methods, section (b). The tube was photographed in ultraviolet light.



(c) Analytical centrifugation

Buoyant density experiments were performed in Beckman model E ultracentrifuges equipped with photoelectric scanners at 44,000 or 44,770 rev./min at 25°C for 24 hr. Neutral and alkaline buoyant densities were calculated from the distances between the M DNA peaks and a crab dAT marker. In these calculations the buoyant density of the marker was taken to be $1.669_6$ g/ml. at neutral pH (Bauer & Vinograd, 1968) and 1.728 g/ml. at high pH. The buoyant density gradient (Vinograd & Hearst, 1962) was used in the calculations.

The band sedimentation velocity experiment was performed in 2.85 M-CsCl, pH 12.5, 20°C, 35,600 rpm and recorded with the photoelectric scanning system in a Beckman model E ultracentrifuge.

(d) Hybridization observed in the electron microscope

The nicked circular M DNA mixture used for the electron microscope hybridization studies was prepared by illuminating a cellulose nitrate SW50 tube containing a single band of freshly prepared closed circular M DNA with a 100W incandescent light bulb at 8 inches for 36 hr. The tube contained 300 $\mu$g/ml. EB and 4.5 M-CsCl, 0.01 M-Tris, pH 7.5, that had been centrifuged at 43 Krpm for 24 hr. The closed DNA had been isolated as a lower band and had been rebanded in a second tube. Approximately 50% of the closed circular M DNA was photochemically nicked as estimated by the ratio of upper to lower band material upon recentrifugation to equilibrium in

CsCl-EB. The DNA from the upper band material (fraction $D_1$) was freed of EB and pelleted as described above. The foregoing procedures were designed to prepare an M DNA sample in which all molecules were nicked at least once, and seldom twice, as discussed in the results.

Approximately 0.05 $\mu$g of M DNA (fraction $D_1$) in 10 $\mu$l. was denatured by dialysis against 95% formamide and 0.01 M-EDTA at pH 8.3 for 1 hr at 24°C, and was reannealed by dialysis for $2\frac{1}{2}$ hr against 50% formamide, 0.20 M-Tris, pH 8.0, 0.02 M-EDTA. The reaction was then quenched by dialysis for 2 hr against 0.1 M-NaCl and 0.01 M-EDTA, pH 7.3. The dialysis mixture was divided into two samples. One sample was adjusted to contain 0.1% cytochrome C and 0.5 M-$NH_4Ac$ at pH 7.0 in a final volume of 50 $\mu$l. This sample was then spread onto 0.25 M-$NH_4Ac$ at pH 7.0. (This procedure will be referred to as the aqueous technique.) The second sample of renatured DNA was adjusted to contain 0.1% cytochrome C, 0.1 M-$NH_4Ac$, 0.01 M-Tris at pH 8.0, and 40% formamide. The latter sample was then spread onto 0.01 M-Tris at pH 8.0 and 10% formamide. (This procedure will be referred to as the formamide technique.) The films were picked up on Parlodion-coated specimen grids and were either stained with uranyl acetate (Davis & Davidson, 1968) or shadowed with platinum-palladium (Pt-Pd).

Nicked circular PM2 viral DNA, prepared by the method of Espejo & Canelo (1968), was a gift from R. Watson. The sample obtained as an upper band from a CsCl-EB density gradient was

lightly nicked and contained on the average about three single-strand scissions per molecule. This was shown in an analytical sedimentation velocity experiment in alkaline CsCl, pH 12.5. Approximately 20% of the DNA was in the form of single-strand circles and approximately 50% as single-strand linears. The rest of the material sedimented as smaller molecular weight linear species. Denaturation, reannealing, and specimen preparation was performed as described above.

Electron microscopy was routinely employed as a check on the form of the M DNA samples. Specimen grids of native duplex M DNA were prepared as described previously (Clayton & Vinograd, 1967). A Philips EM 300 electron microscope equipped with an auxiliary Plumbicon TV system was used.

(e) Reagents, enzymes, and marker DNA

Optical grade CsCl from the Harshaw Chemical Company, Cleveland, Ohio, and ethidium bromide from Boots Pure Drug Co., Ltd., Nottingham, England, were used without further purification. All other chemicals were reagent grade. DNase I and RNase A used in the mitochondria isolation were purchased from the Sigma Chemical Company, St. Louis, Missouri. Cytochrome C (Lot 45461) was purchased from Calbiochem, Los Angeles, California. Crab dAT isolated by the $Cs_2SO_4$-$HgCl_2$ method (Davidson et al., 1965) from Cancer antennarius sperm was a gift from R. Hyman.

### (f) Measurement of pH

A Beckman Research model pH meter with an E. H. Sargent combination small-probe glass electrode was used. The pH meter was standardized at 24° C with Beckman saturated $Ca(OH)_2$ buffer at pH 12.63 before the pH determinations of alkaline CsCl solutions were performed. The buffer was re-read after a determination. The drift did not exceed ± 0.05 pH unit.

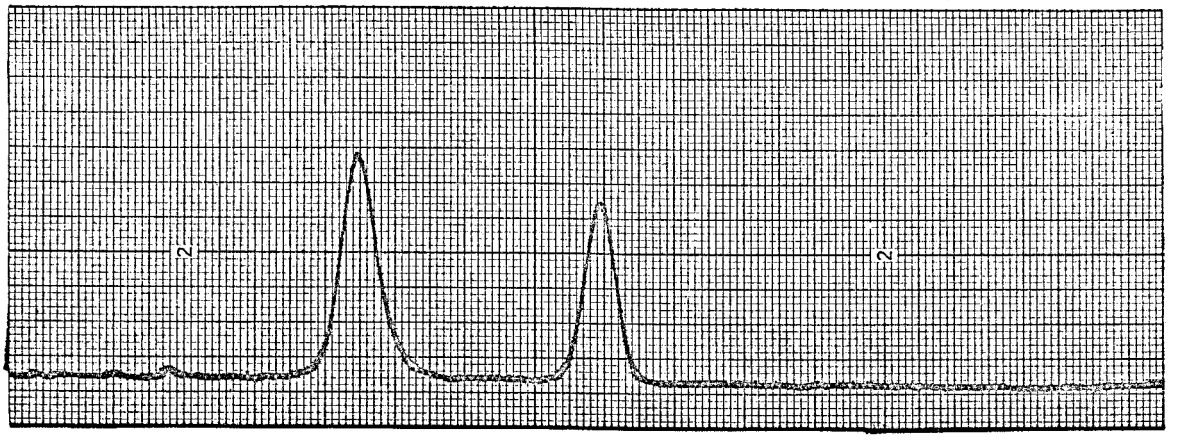
## 3. Results

### (a) Neutral buoyant densities of circular dimer and monomer M DNA

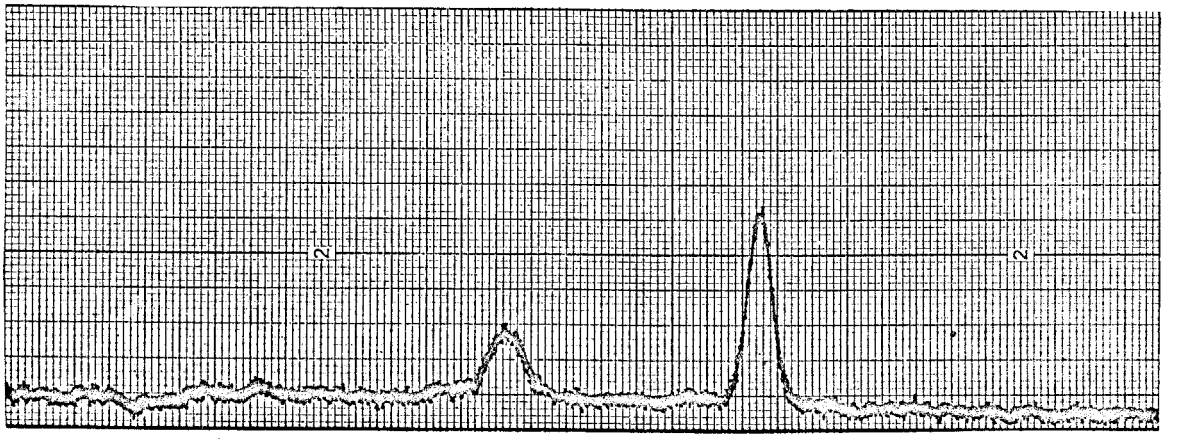
It is well known that there is usually a correspondence between the buoyant density of DNA and its base composition (Schildkraut, Marmur & Doty, 1962). We have previously reported a buoyant density for leukocyte M DNA of 1.700 g/ml. and a corresponding calculated base composition of 46 mole percent guanine-cytosine. This analysis was performed on a mixture of M DNA forms. Figure 3 presents scans of absorbance profiles of analytical cells containing the separated circular dimers and monomers. A marker DNA, crab dAT, was present in each experiment. The calculated buoyant densities of the monomer and the circular dimer species, 1.6996 ± 0.0005 g/ml., are in good agreement with the value reported earlier for the mixture. The overall base compositions of the circular dimer and monomer are, therefore, identical within the experimental error

Figure 3. Photoelectric scans of (a) purified monomeric and (b) purified circular dimeric leukemic leukocyte M DNA with a crab dAT marker in neutral buoyant cesium chloride, 25°C, 44, 770 and 44, 000 rpm, respectively. The field is directed to the right. The light band contains dAT and the dense band M DNA.

a)



b)



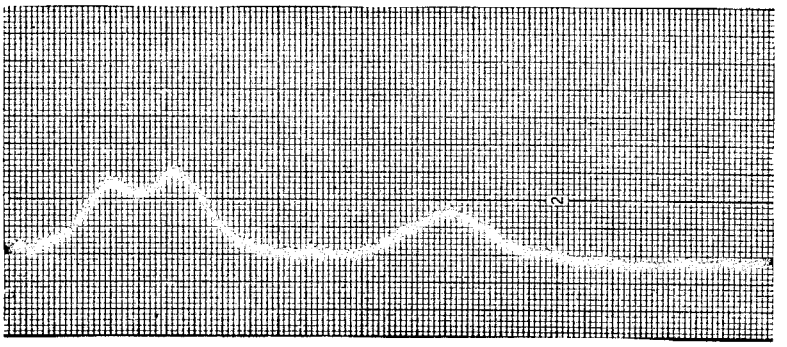
limits of ± 1.0 mole percent guanine-cytosine (GC). A difference of 0.001 g/ml. in buoyant density corresponds to a difference of 1 mole percent GC (Schildkraut, Marmur & Doty, 1962).

(b) Alkaline buoyant densities of the complementary strands in circular dimer and monomer M DNA

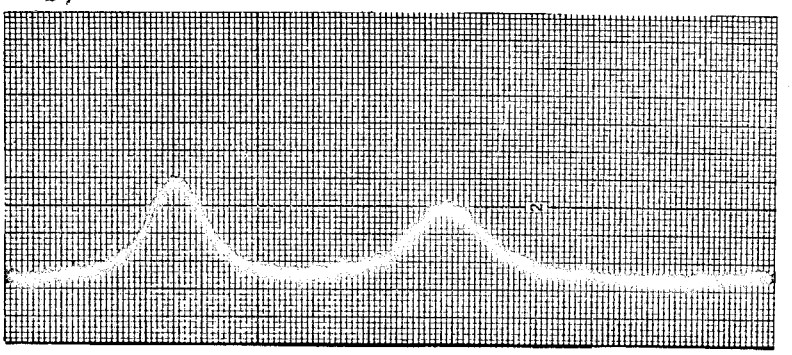
Centrifugation of nicked circular or linear DNA in buoyant alkaline (pH 12.5) CsCl usually gives a unimodal distribution of DNA at a higher density than the native duplex, due to the titration of guanine and thymine residues and subsequent loss of the duplex structure (Vinograd, Morris, Davidson & Dove, 1963). Human M DNA in such experiments forms a bimodal distribution. Each band contains one of the two M DNA complementary strands (Corneo, Zardi & Polli, 1968). The difference in buoyant density of the two strands is thought to be principally the result of a difference in the guanine plus thymine content of each strand. Scans of the buoyant profiles of purified circular dimers (fraction $B_2$) and purified monomers (fraction $B_1$) are presented in Figure 4. The buoyant densities of the light strand and heavy strand of the circular dimer were calculated from results of an experiment in which crab dAT was added as a density marker (Fig. 4(a)). The light and heavy strands are 1.738 and 1.779 g/ml., respectively. The separation between the complements is the same, 0.041 g/ml., in the dimer samples with and without the marker (Figs. 4(a) and (b)). The separation between complementary strands in the preparation of monomers is also 0.041 g/ml. (Fig. 4(c)). In Figure 4(d) the marker obscured the small amount of light monomer

Figure 4. Photoelectric scans of leukemic leukocyte M DNA in buoyant alkaline cesium chloride, pH 12.5, 0.05 M-$K_3PO_4$, 25°C. The field is directed to the right. (a) The least dense band is crab dAT. The band at intermediate density is the light strand of purified circular dimer M DNA (fraction $B_2$). The dense band is the heavy strand of purified circular dimer M DNA. (b) The light strand and heavy strands of purified circular dimer M DNA (fraction $B_2$). (c) The light and heavy strands of purified monomers (fraction $B_1$). (d) A mixture of purified monomers (fraction $B_1$) and crab dAT. The M DNA light strand is observed as a shoulder on the dense side of the marker. The heavy strand forms a broad band near the center of the pattern (↓). All of the experiments were at 44, 770 rpm, except (c), which was at 44, 000 rpm.

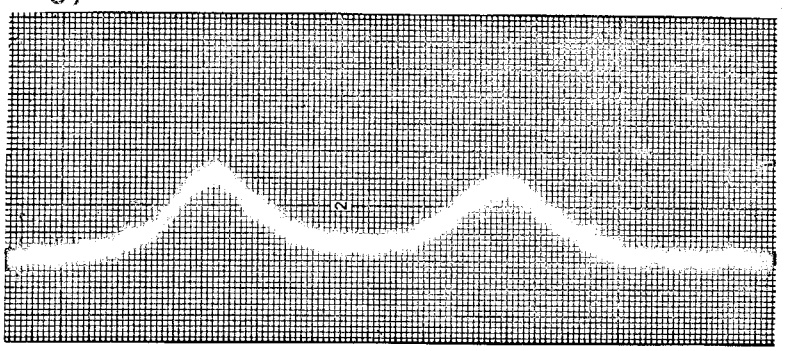
a)



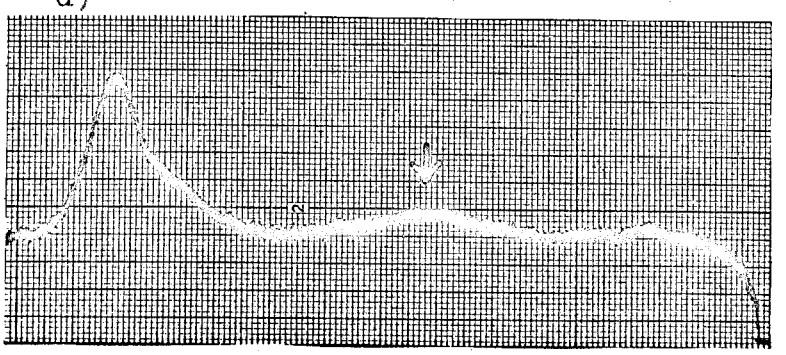
b)



c)



d)



strand present. The maximum in the center of the diagram corresponds to 1.779 g/ml. for the heavy monomer strand. The buoyant density of the light monomer strand, 1.738 g/ml., was calculated from the separation between strands observed in Figure 4(c) and from the value for the heavy strand in Figure 4(d). We conclude that the base compositions of the corresponding circular dimer and monomer complementary strands are likely to be very similar, if not identical.

It might be argued that the presence of the light and heavy bands (Figs. 4(a) and (b)) in the buoyant pattern obtained with dimer DNA at high pH and the absence of bands midway between them rules out the head-to-head structure (Fig. 1(b)) for the dimer. Each of the dimer strands in this model, if intact, should have a buoyant density close to the mean of the heavy and light monomer strands. Unfortunately, the mitochondrial DNA strands cannot be assumed to be full-length in these experiments because of the well known sensitivity of M DNA to alkali. The molecular weight, calculated from the band width at the six-tenths height, is about one-third that expected for full-length strands. Such fragmented dimer strands, if derived from the head-to-head structure, should form skewed bands at the positions of heavy and light monomer strands and elevate the base line between the bands. The bands, however, are fairly symmetrical and there is little material between. It can also be argued that material with intermediate buoyant density would be entirely absent if the two bonds joining the monomeric genomes were

were specially alkali sensitive, a possibility which has not been ruled out. The electron microscope studies (Materials and Methods, section (d)), performed with intact strands, show that the model in Figure 1(b) cannot be correct. The centrifuge results cited above are consistent with this conclusion.

### (c) The reannealing of light monomeric and heavy dimeric M DNA complements

A direct analysis of the ability of circular dimer single strands to hybridize with monomer single strands can be obtained by examining the buoyant density patterns of artificial mixtures of heavy dimer and light monomer strands that have been subjected to reannealing conditions. The heavy strand of purified circular dimers and the light strand of purified monomers were isolated by alkaline buoyant density centrifugation in the preparative ultracentrifuge.

Approximately 0.5 $\mu$g of nicked circular dimer (fraction $B_2$) and 0.5 $\mu$g of nicked monomer (fraction $B_1$) M DNA were placed in separate alkaline buoyant CsCl solutions, 1.760 g/ml., 0.05 M-$K_3PO_4$, pH 12.5, in Polyallomer tubes. In addition, approximately 0.5 $\mu$g of previously isolated light strands of [$^3$H]HeLa M DNA was mixed with the circular dimer sample and approximately 0.05 $\mu$g of heavy strands of [$^3$H]HeLa M DNA was mixed with the monomer sample. The specific activity of the HeLa DNA strands was approximately $10^5$ cpm/$\mu$g. The corresponding buoyant densities of

complementary strands of HeLa M DNA and leukemic leukocyte M DNA in alkali are known to be similar (R. L. Hallberg, private communication). The alkaline gradients were centrifuged for 65 hours at 20°C and at 31.5 Krpm in an SW50.1 rotor. Forty-$\mu$l. fractions were collected, and 5-$\mu$l. aliquots were counted to locate the position of the [$^3$H]HeLa M DNA strands (Fig. 5). The fractions containing the separate strands of the circular dimer and monomer leukocyte M DNA were appropriately pooled (Fig. 5), and either cross-hybridized or "self-annealed" by dialysis against 50% formamide, 0.20 M-Tris, pH 7.5, 0.01 EDTA for ten days. The samples were dialyzed against 0.50 M-NaCl, 0.01 Tris, pH 7.5, and 0.005 EDTA to remove the formamide and then pelleted for twelve hours at 40,000 rev./min in an SW50.1 rotor as described earlier. The pellet was suspended in concentrated CsCl solution, 0.01 M-Tris HCl, pH 7.5, and examined at buoyant equilibrium in the analytical ultracentrifuge.

The resultant hybrid molecules formed a single sharp band with a buoyant density of 1.700 g/ml. (Fig. 6(a)). This value is the same as that obtained for both the monomeric and dimeric native duplex M DNA's. We conclude that a high molecular weight duplex species has formed. This species contains heavy strands from the circular dimer and light strands from the monomer. No banded material was observed at positions for neutralized single strands of light monomer, 1.707 g/ml. (Fig. 6(b)), or heavy dimer.

The self-annealed light strands of the circular dimer were examined as a check for contamination and for any evidence of self-annealing (Fig. 6(b)). The material formed a band with a buoyant

Figure 5. [$^3$H]HeLa light and heavy M DNA complements were used to detect the position of light and heavy M DNA complements in leukocyte M DNA, as described in the text. Figure 5(a) is a radioactive pattern of HeLa M DNA complements obtained by R. L. Hallberg in this laboratory using the conditions employed in (b) and (c). The peaks have a separation of 13 fractions. Since the monomer sample, (b), has a band maximum for the [$^3$H]HeLa heavy strand at fraction 20, we expect the band maximum for the light leukocyte monomer strand to be located at fraction 33. Similarly for the circular dimer sample, (c), the band maximum for the [$^3$H]HeLa light strand was found at fraction 32. The band maximum for the heavy dimer strand is expected to be at fraction 19. Therefore, fractions 30 through 40 in (b) were pooled and used as the monomer light-strand sample and fractions 12 through 22 in (c) were pooled and used as the dimer heavy-strand sample. The light dimer strand (fractions 24 through 40 in (c)) was isolated and used as a control.

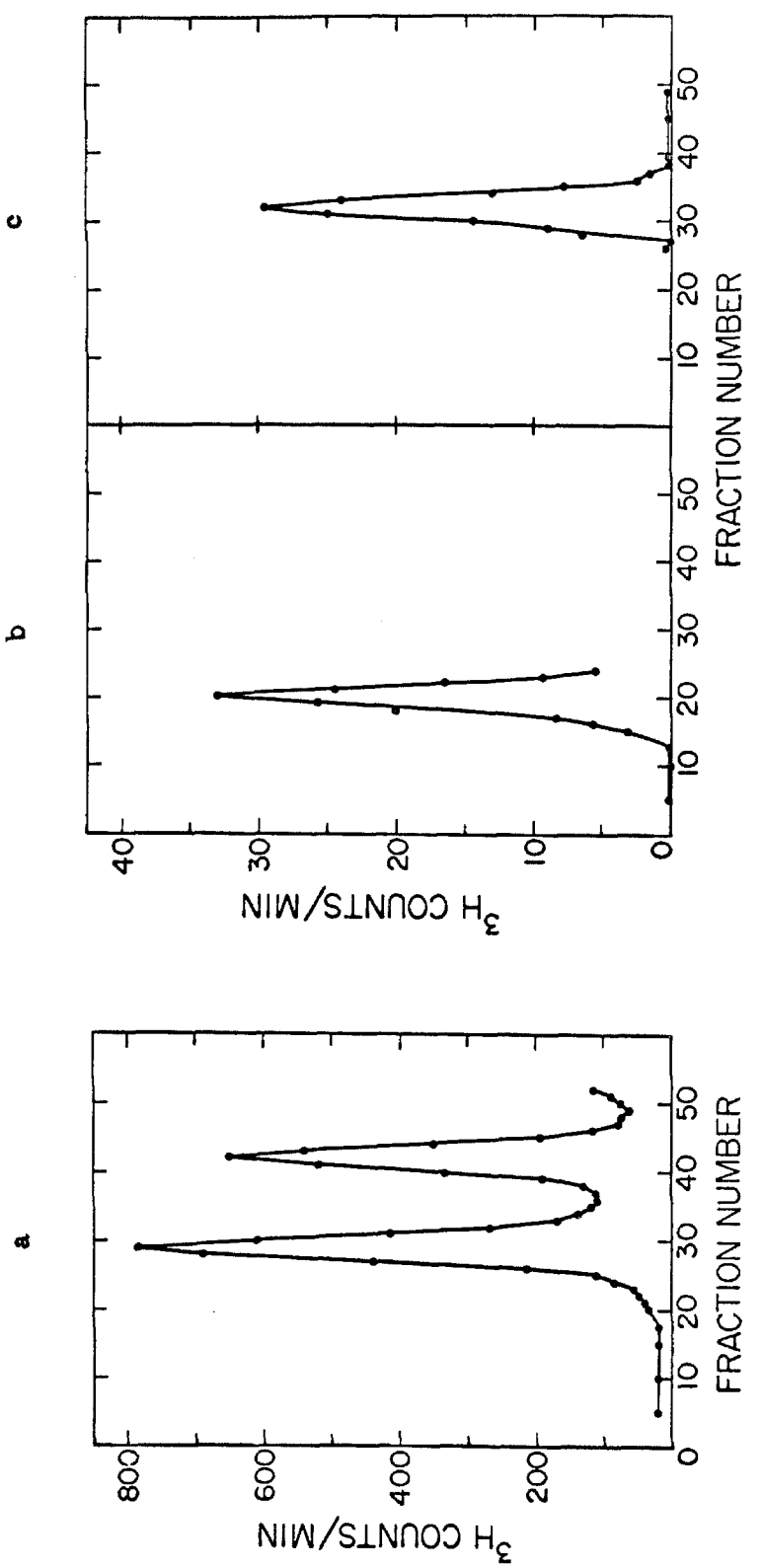
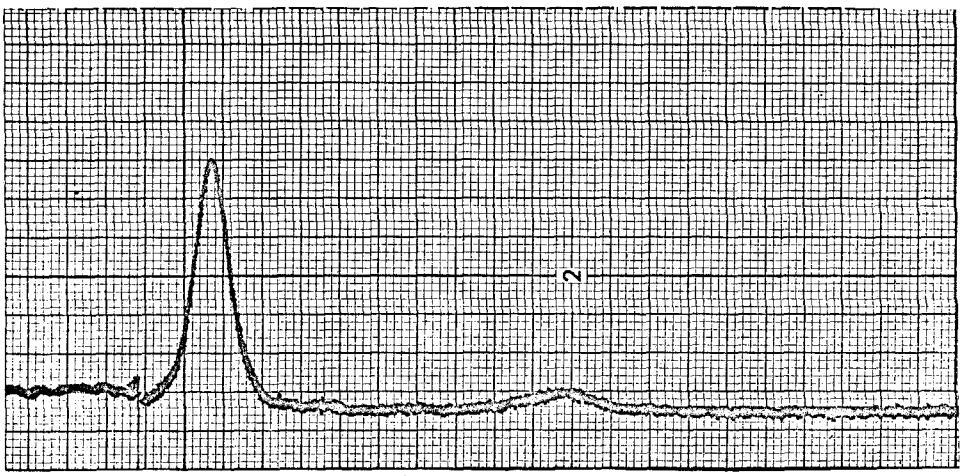

a
b
c
$^3$H COUNTS/MIN
FRACTION NUMBER
0
200
400
600
800
10
20
30
40
50

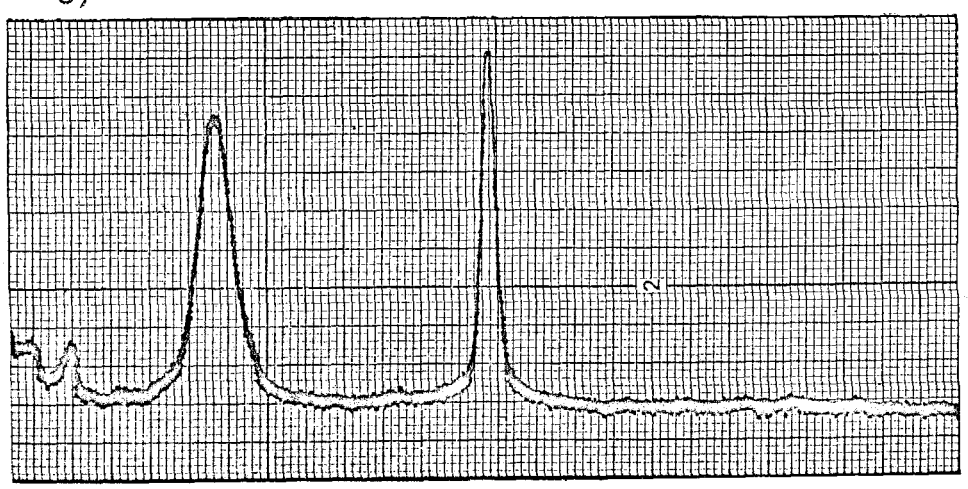
Figure 6. Photoelectric scans of annealed M DNA in neutral buoyant cesium chloride at 44, 770 rpm, 25°C. The field is directed to the right. The light band in each case is crab dAT added as a marker. (a) Annealed mixture of approximately equal amounts of heavy dimer strand and light monomer strand plus crab dAT. (b) Self-annealed light dimer strand plus crab dAT. (c) Annealed mixture of denatured monomer and dimer M DNA plus crab dAT. The mixture contained 88% dimer and 12% monomer.

a)


b)




c)



density of 1.707 g/ml. The light strand of HeLa M DNA forms a band at neutral pH with the same buoyant density (R. L. Hallberg, private communication). No contaminating heavy single-stranded DNA or duplex DNA was detected. The sample of 88% dimers and 12% monomers from the alkaline buoyant density experiment in Figure 4(b) was neutralized, annealed, and centrifuged in a neutral buoyant CsCl gradient. The hypersharp band with a buoyant density of 1.700 g/ml. (Fig. 6(c)) indicates that a mixture of annealed dimer and monomer strands regains a duplex structure analogous to the hybrid molecules in Figure 6(a).

In view of the similarity of the neutral and alkaline buoyant behavior between leukocyte and HeLa M DNA, we take 1.707 g/ml. as the buoyant density of the neutralized light strand (Fig. 6(b)) and 1.726 g/ml. as the buoyant density of the heavy strand (R. L. Hallberg, private communication). The mean, 1.717 g/ml., is 0.017 g/ml. higher than the value for the native duplex. The hybrid structure with the same buoyant density must, therefore, within experimental error, consist of pure duplex. Regions of nonhomology or imperfect annealing in the concatenate should have raised the buoyant density of the hybrid. We estimate the uncertainty in $\Delta\Theta$ in these experiments to be about $\pm 0.001$ g/ml. Based on this uncertainty and assuming linearity between the fractional duplex content and the fractional buoyant increment (Rownd, Lanyi & Doty, 1961), we estimate that at least 90% of the mass of the buoyant species is in a duplex structure. We conclude, therefore, that any heterologous regions, if present,

cannot exceed 10% of the total in the monomeric and dimeric M DNA forms. A summary of the sedimentation properties of circular monomeric and dimeric forms is given in Table 1.

(d) Electron microscopy of reannealed M DNA

The apparent identity of base compositions of the circular dimer and monomer forms of leukemic leukocyte M DNA and the existence of at least 90% homology between the two forms suggests that the circular dimer is essentially a double-sized copy of the monomer genome. To test this hypothesis further, we have employed the electron microscope technique for the detection of heterologous regions in reannealed DNA.

Heteroduplex molecules containing one strand from one M DNA molecule and the complementary strand from another were prepared by renaturing a mixture of denatured (fully strand-dissociated) monomers and dimers. Prior to denaturation at least one single-strand scission was introduced into each DNA molecule. Since a second scission in a strand of the duplex results in fragmented single strands which interfere with the analysis, this study was carried out with a mixture of lightly nicked molecules (fraction $D_1$). The composition of this fraction is given in Table 2. All of the duplex forms contained at least one scission, as indicated in Materials and Methods, section (d). We have calculated with the aid of the Poisson relation that at least 88% of the monomeric and

TABLE 1

Sedimentation properties of the circular monomeric and dimeric forms of human leukemic leukocyte M DNA

| | Monomer | Dimer |
|---|---|---|
| $\underline{s}^0_{20,w}$, closed, Svedbergs | — | 51.6 ± 0.7* |
| $\underline{s}^0_{20,w}$, nicked, " | 25.6 ± 0.4† | 33.1 ± 0.5‡ |
| $\underline{s}^0_{20,obs,alk.}$, closed, " | 80** | 112** |
| $\Theta$, g/ml., duplex, neutral | 1.700 | 1.700 |
| $\Theta$, " , light strand, alk. | 1.738 | 1.738 |
| $\Theta$, " , heavy strand, " | 1.779 | 1.779 |
| $\Theta$, " , light strand, neutral | — | 1.707 |

* Clayton & Vinograd (1967). †Previously unpublished data. ‡Hudson, Clayton & Vinograd (1968). The sedimentation solvent was 2.85 M-CsCl at 20°C. The sedimentation coefficients have been fully corrected to standard conditions and are expressed as values for the sodium form.

** Clayton & Vinograd (1967). These sedimentation coefficients are the observed uncorrected values in 2.85 M-CsCl, 0.05 M-$K_3PO_4$, pH 12.5, 20°C.

TABLE 2

Electron microscope frequency analysis of M DNA forms before denaturation and after annealing

| Form | Before denaturation (%)* | After annealing (%) |
|---|---|---|
| Circular monomers | 71 | 75 |
| Circular dimers | 24 ± 3 | 4 ± 2 |
| Catenated dimers | 3.5 ± 1 | 9.5 ± 2† |
| Tailed molecules | 0 | 3 ± 1 |
| Linear duplexes | 1.5 ± 1 | 1 ± 0.9 |
| Ambiguous molecules | 0 | 8.5 ± 2 |
| Total no. of molecules classified | 800 | 800 |

* The error values indicate the interval which contains the true mean at a level of confidence of 95%.

† Most of these molecules are considered to be fused dimers, as described in the text.

80% of the circular dimeric strands were present during the re-annealing experiment as full-length linear or circular single strands. The remainder must have been present as linear fragments of varying size.

In order to avoid introducing extra single-strand scissions, which occur at high pH or at elevated temperature, the M DNA was denatured and renatured with formamide. A micro-procedure, using only 0.05 $\mu$g of DNA, was employed. The renatured DNA sample was mounted on a specimen grid for electron microscopy by the formamide technique (Davis, Simon & Davidson, 1969). The formamide was incorporated into the mounting solution to melt out the random base interactions of single-stranded DNA. The heterologous regions are more readily detected with this mounting procedure than with the aqueous technique. Several kinds of duplex DNA's, but little single-stranded material, were observed. Plate I shows a typical field as viewed in the electron microscope. All duplex molecules were examined for heterologous regions (Figs. 7(a) and (b)). Several hundred molecules were studied, but no heterologous regions were seen in any of the duplex forms. We conclude that the sequence similarity between the M DNA's is so great that no detectable heterologous regions occur. It has been our experience that deletions, additions, or substitutions as small as 50 to 100 base pairs can be detected by this procedure.

In a separate experiment to test if the procedure used here does indeed denature the DNA and allow for strand dissociation, the

Plate I. Typical field of renatured M DNA as viewed in the electron microscope. Three circular duplex molecules of monomer length, a circular duplex molecule of two monomer lengths, a single-stranded circular monomer molecule (upper-right corner), and a linear duplex molecule of slightly less than monomer length and with single-stranded ends (lower-right corner) are shown. The DNA was mounted on specimen grids by the formamide technique and shadowed with Pt-Pd.

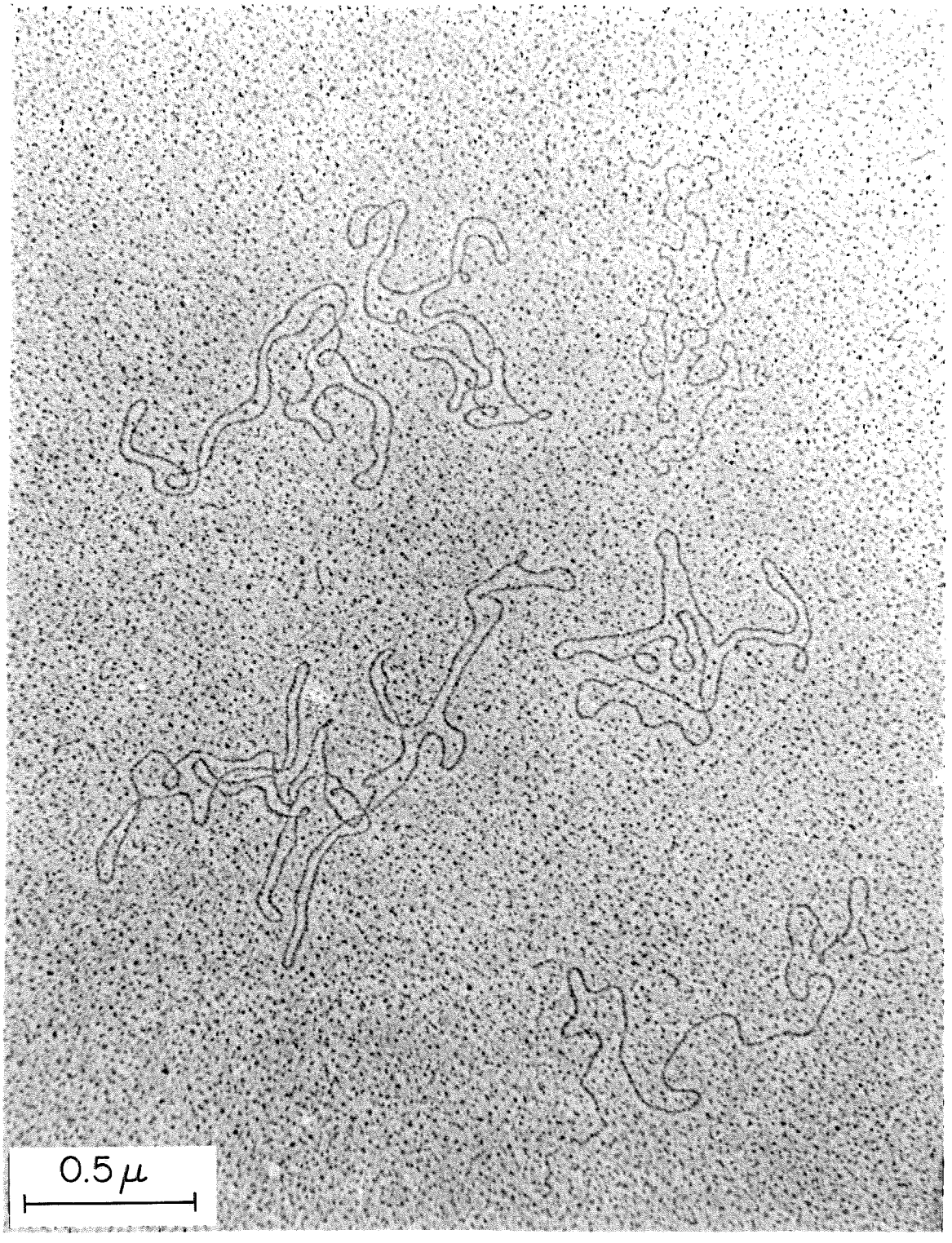
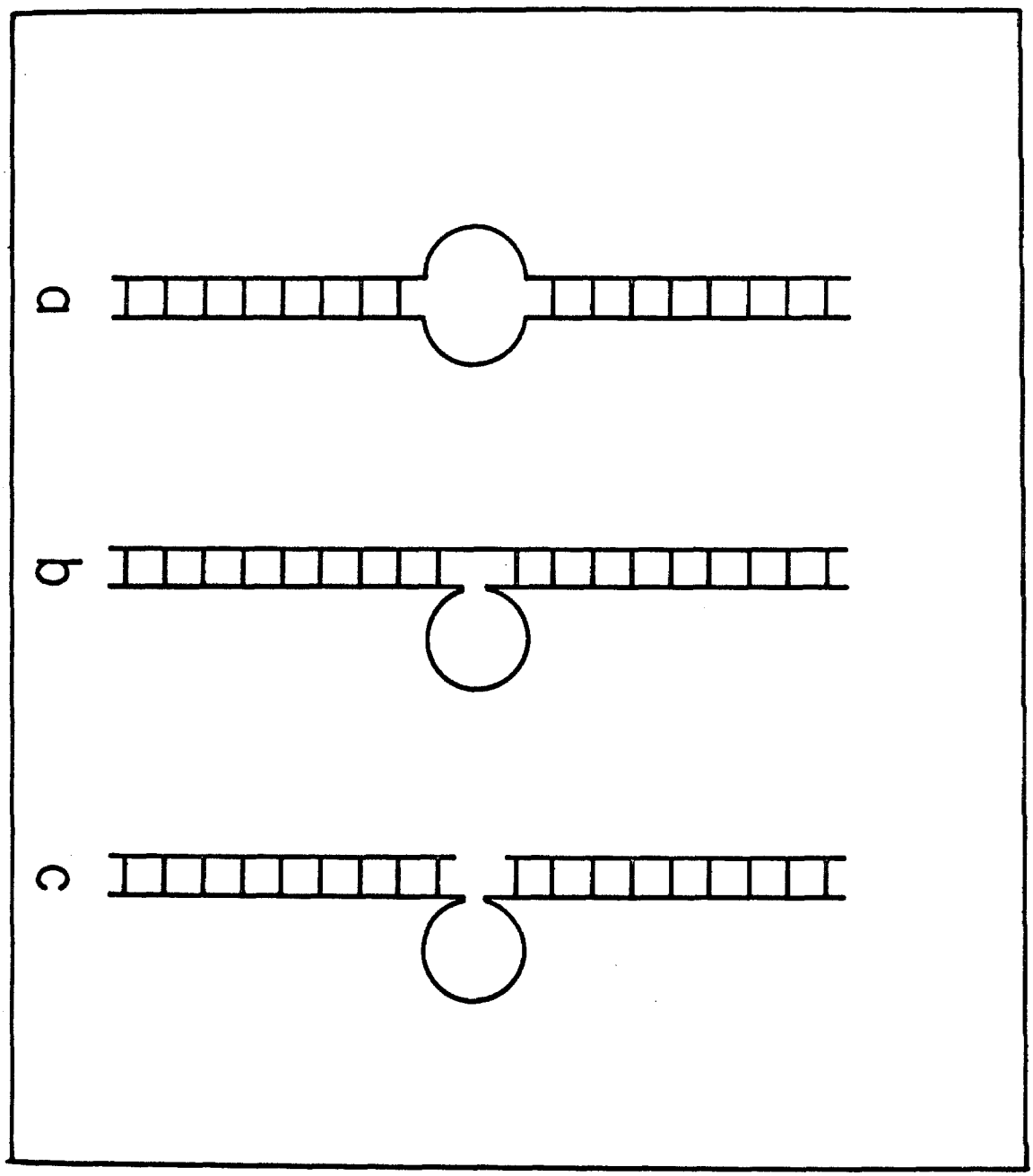

0.5 μ

Figure 7. Representations of three ways of forming a nonduplex region. (a) A region of nonhomology. (b) An insertion or deletion. (c) A gap region.


a
b
c

reannealing step was omitted after denaturation. The resulting sample contained only 6% of the DNA molecules in the form of circular duplex monomers. These were presumably renatured molecules that formed during the quenching step. The remainder were identified from their appearance as single strands. We conclude, therefore, that the complementary strands in the mixture experiment had separated from each other upon denaturation and had reformed duplexes on reannealing.

In order to visualize the topology of the duplex forms more readily, the reannealed mixture of DNA's was also mounted on specimen grids for electron microscopy by the aqueous technique. The formamide technique often causes acute tangling of the DNA and makes topological identification difficult. The duplex molecules were scored as monomers, dimers, etc. Plate II shows the three major classes of duplex molecules found in this study. It should be recalled that single strands of DNA, when mounted by the aqueous techniques, collapse into what appears to be a "bush" due to random base interactions.

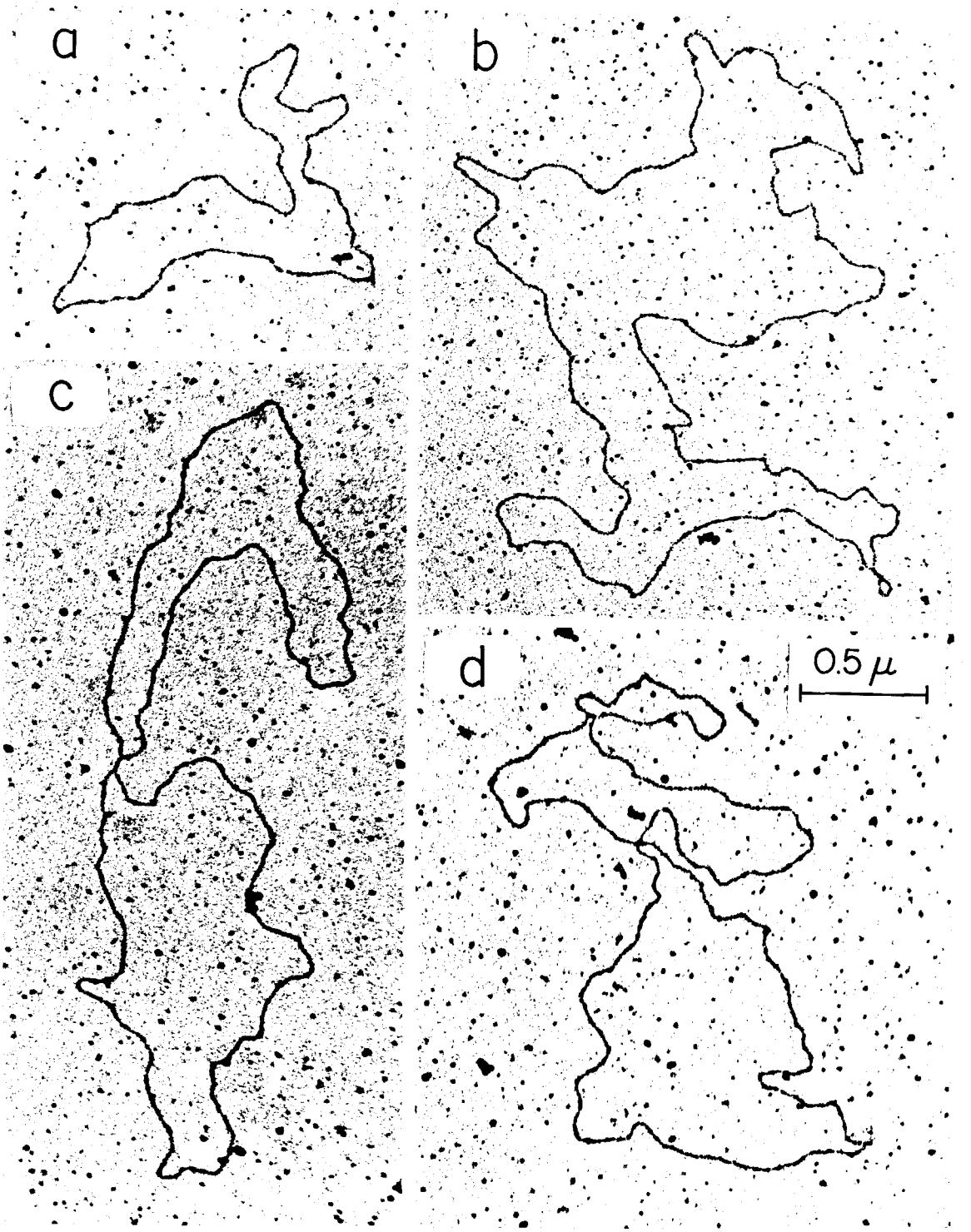
If no hybridization between monomer and dimer strands had occurred, we should have found monomer and dimer circular duplexes and catenanes in approximately the same proportion as in the original sample. We should also have found some linear duplexes formed from originally circular duplexes in which both strands were nicked. Most of the latter kind of molecules should have cyclized in these experiments. We did observe, however,

Plate II. Major classes of renatured M DNA. (The DNA was mounted on specimen grids by the aqueous technique and stained with uranyl acetate.)

(a) Renatured circular monomer.

(b) Renatured circular dimer.

(c) and (d) Renatured "figure 8" molecules. Such molecules are considered to be fused dimers, as explained in the text.

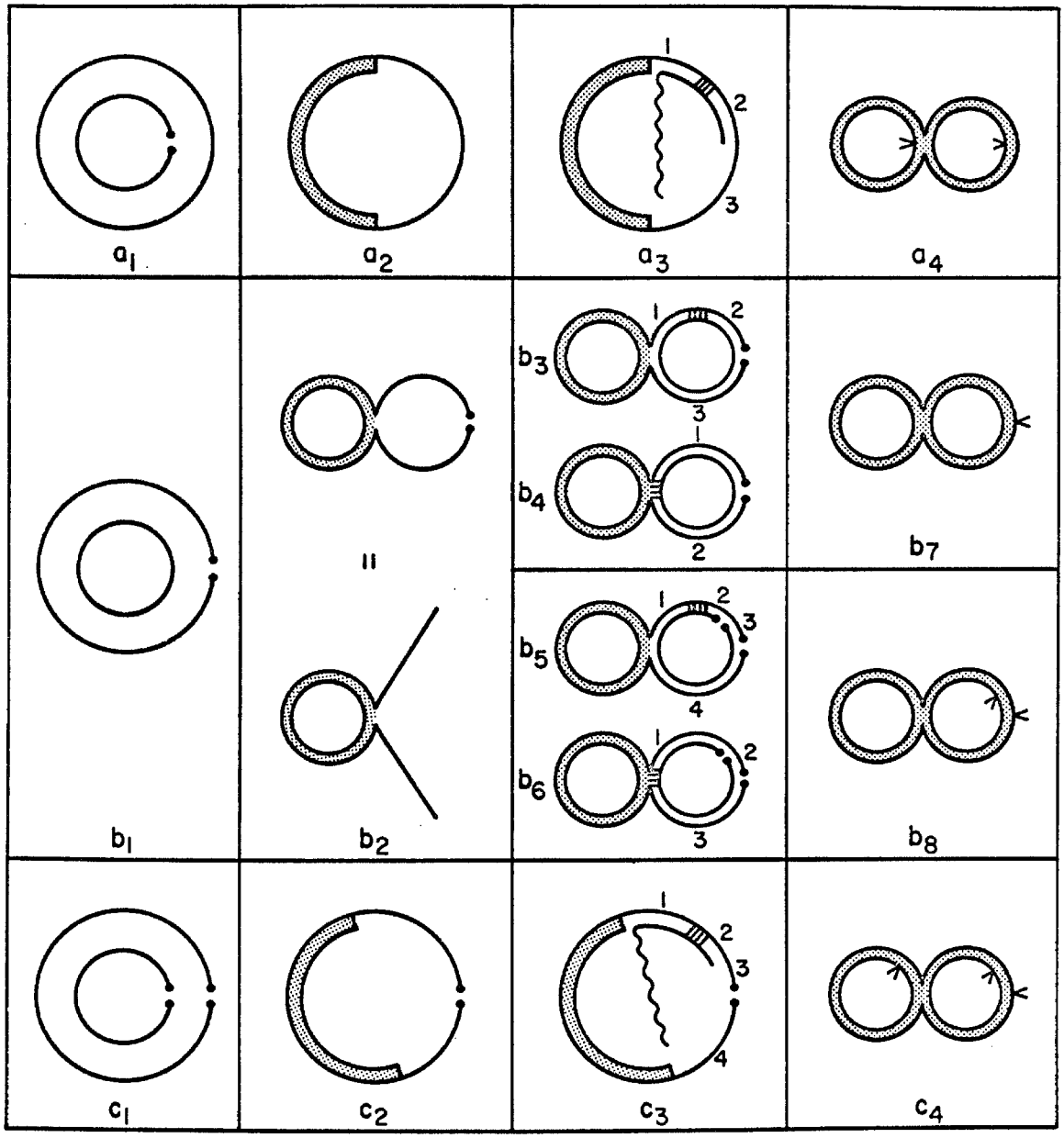

a
b
c
d
0.5 μ

that the frequency of the circular dimer (Plate II(b)) decreased from 24 to 4%, and the frequency of "figure 8" molecules (Plates II(c) and (d)) increased from 3 to 9.5% (Table 2). The frequency of monomers (Plate II(a)) remained essentially the same: 71 and 75%. Circular monomer duplexes with one or two double- or single-stranded tails were also observed at a frequency of 3% (Plate IV).

The absence of 5-$\mu$ linear duplexes is regarded as strong evidence for the head-to-tail structure of the dimer (Fig. 1). The head-to-head dimer should have preferentially reannealed by a monomolecular mechanism to form uniform linear duplexes of 5-$\mu$ contour lengths. The two hairpin regions would be too small to appear as bushes. No 5-$\mu$ linear duplexes were observed. Also, no 5-$\mu$ linear duplex molecules were present in the sample of denatured DNA that was slowly quenched. The head-to-head dimer should have annealed under these conditions.

We consider next the various pathways for the formation of heteroduplex molecules from full-length single strands. Figure 8 presents schematically some possible modes of formation of heteroduplexes from the four main types of full-length single strands present in the mixture during annealing. These may pair in three ways. Reannealing between circular complements is regarded as unlikely and is not considered. Heteroduplex formation between a linear monomer and a circular dimer is shown in Figure 8($a_2$). When a second linear monomer anneals with ($a_2$), an intermediate structure, ($a_3$), is formed. Regions 1 and 2 of the circularly permuted linear

Figure 8. Representation of various modes of formation of heteroduplexes containing full-length monomer strands and one full-length dimer strand. These duplexes are called fused dimers and are shown in the fourth column. The paired strands in the first column anneal to form the structures containing two single strands in the second column. These nucleate with a third single strand as shown in the third column and anneal to form fused dimers. The shaded areas represent hydrogen-bonded duplex regions. The unshaded annuli represent nonduplex regions. The nucleation sites are indicated by transverse lines. The carats (v) represent single-strand scissions. A detailed description of the winding that occurs in the formation of the fused dimer is given in the text.


a1
a2
1
2
3
a3
a4
b1
b2
b3
1
2
3
b4
1
2
b7
b5
1
2
3
4
b6
1
2
3
b8
c1
c2
1
2
3
4
c3
c4

strand wind around the complementary regions in the circular dimer strand to form a hydrogen-bonded duplex, and leave an unannealed tail complementary to region 3. This tail winds around the dimer strand after a second nucleation anywhere in region 3. There are no topological restrictions in this process because of the presence of swivels at both ends of this region. The duplex, ($a_4$), contains two single-strand scissions at the positions indicated by the carats (v). Such "figure 8" molecules have the appearance of catenanes in electron micrographs. We shall refer to these molecules as fused dimers.

The intermediate (Fig. 8($b_2$), made up of a linear dimer and circular monomer strand could appear with varying tail lengths, depending upon the site of nucleation. We illustrate the formation of fused dimer molecules from an intermediate with equal-size single-stranded tails. Four illustrations of the modes of formation of fused dimers are given in Figures 8($b_3$) through ($b_6$).

Figure 8($b_3$) illustrates the nucleation of a monomer circle with a site on a single-stranded tail. Duplex formation in region 2 can occur by winding of this chain-end around the circular strand. Duplex formation in region 1 can occur by rotation of the partially duplex circular monomer around the linear strand in region 1. The linear tail in region 3 can then reanneal to complete the formation of the fused dimer, ($b_7$), with one single-strand scission. Figure 8($b_4$) illustrates nucleation immediately adjacent to the junction region. Winding of each of the tails then results in the formation of the fused dimer.

Figures 8($b_5$) and ($b_6$) illustrate intermediates that arise when similarly permuted linear strands nucleate with single-stranded regions in ($b_2$). Winding in ($b_5$) occurs in the regions 1, 2, and 4, and leaves cohesive ends in region 3 unreacted. A cyclization reaction, such as described by Thomas & MacHattie (1964) for the circularly permuted strands of T2 DNA and by Hershey & Burgi (1965) for the cohesive ends of lambda DNA, then occurs and leads to a fused dimer, ($b_8$). Fused dimers can also be formed as in Figure 8(c) from a linear dimer and two linear monomer strands.

The results of careful inspection and measurement of twenty "figure 8" molecules are presented in Table 3; two examples are shown in Plates II(c) and (d). Seventeen of these molecules are two-ring systems in which the individual lengths of the monomers are within one standard deviation of the mean length. Neither bushes nor tails were observed in this group. Such molecules must have been formed as described above from heterologous strands, unless they were derived from the small number of catenanes (3%) originally present. These catenanes would have formed pairs of interlocked single-stranded rings upon denaturation and could have reannealed with complementary linear monomer strands. If catenanes alone were responsible for the formation of "figure 8" forms, we should have expected to find them at approximately the original frequency. Instead, such molecules accounted for 10% of the duplex forms. Most of the "figure 8" molecules, therefore, have been formed by reannealing of monomer strands with dimer strands. This observation gives

TABLE 3

Normalized contour lengths of annealed human leukemic leukocyte M DNA*

| Monomer | Monomer circles in fused dimers | |
|---|---|---|
| 1.08 | 0.96, 1.01 | 1.04, 1.01 |
| 1.01 | 1.01, 1.01 | 0.99, 0.87‡ |
| 0.99 | 1.00, 0.96 | 1.03, 0.99 |
| 0.98 | 0.98, 0.98 | 0.96, 0.96 |
| 1.02 | 0.98, 1.00 | 0.91, 0.91 |
| 1.00 | 0.99, 0.98 | 0.99, 0.83‡ |
| 1.00 | 1.07, 1.08 | 0.99, 0.99 |
| 1.01 | 1.10, 1.10 | 0.98, 0.65‡ |
| 1.03 | 1.01, 1.03 | 0.99, 1.01 |
| 0.98† | 1.09, 1.07 | 1.05, 1.05 |
| 0.94† | | |
| 1.00† | | |
| 0.99† | | |
| 0.98† | | |
| 1.00 ± 0.031 S.D. | 1.01 ± 0.056 S.D. | |

* The contour lengths were normalized with the mean value of the contour length of the simple monomers.

† Monomer with single-stranded tail.

‡ Circle with bush region. These lengths were not included in the determination of the mean length.

support to the finding that the monomeric and dimeric forms hybridize.

Three of the twenty "figure 8" molecules appeared to contain a single bush in one circular duplex. Such bushes correspond to gap regions which result from the incorporation of short linear dimer or monomer strands into the hybrid (Fig. 7(c)). The contour lengths of the three duplex circles containing bushes were significantly shortened to 2.9, 3.7, and 3.9 $\mu$.

The "figure 8" molecules shown in Plates II(b) and (c) are either renatured catenanes or fused dimers. Since fused dimers are heteroduplexes between a dimer and two monomer strands, it is important to specifically examine these molecules for heterologous regions. Although it is difficult to positively identify a "figure 8" molecule when mounted in the presence of formamide, all possible candidates were carefully studied. Plate III shows a renatured monomer, a renatured dimer, and two "figure 8" molecules. Again, no heterologous regions were found.

So far, we have considered only those renatured molecules which were formed from intact single strands. As previously stated, these experiments were carried out with a mixture of lightly nicked molecules. Approximately 12% of the monomeric strands and 20% of the dimeric strands contained two single-strand scissions and were, therefore, present as fragmented single strands. These strands will, of course, also renature with other whole and fragmented strands and form incomplete duplexes with gaps as shown in Figure 7(c) and Plates IV(d) and (e). The incomplete duplexes, unlike the complete

Plate III. Search for heterologous regions. (The DNA was mounted on specimen grids by the formamide technique and shadowed with Pt-Pd.)

(a) Renatured circular monomer and circular dimer. No heterologous regions can be seen.

(b) and (c) Renatured "figure 8" molecules considered to be fused dimers, as explained in the text. No heterologous regions can be seen.

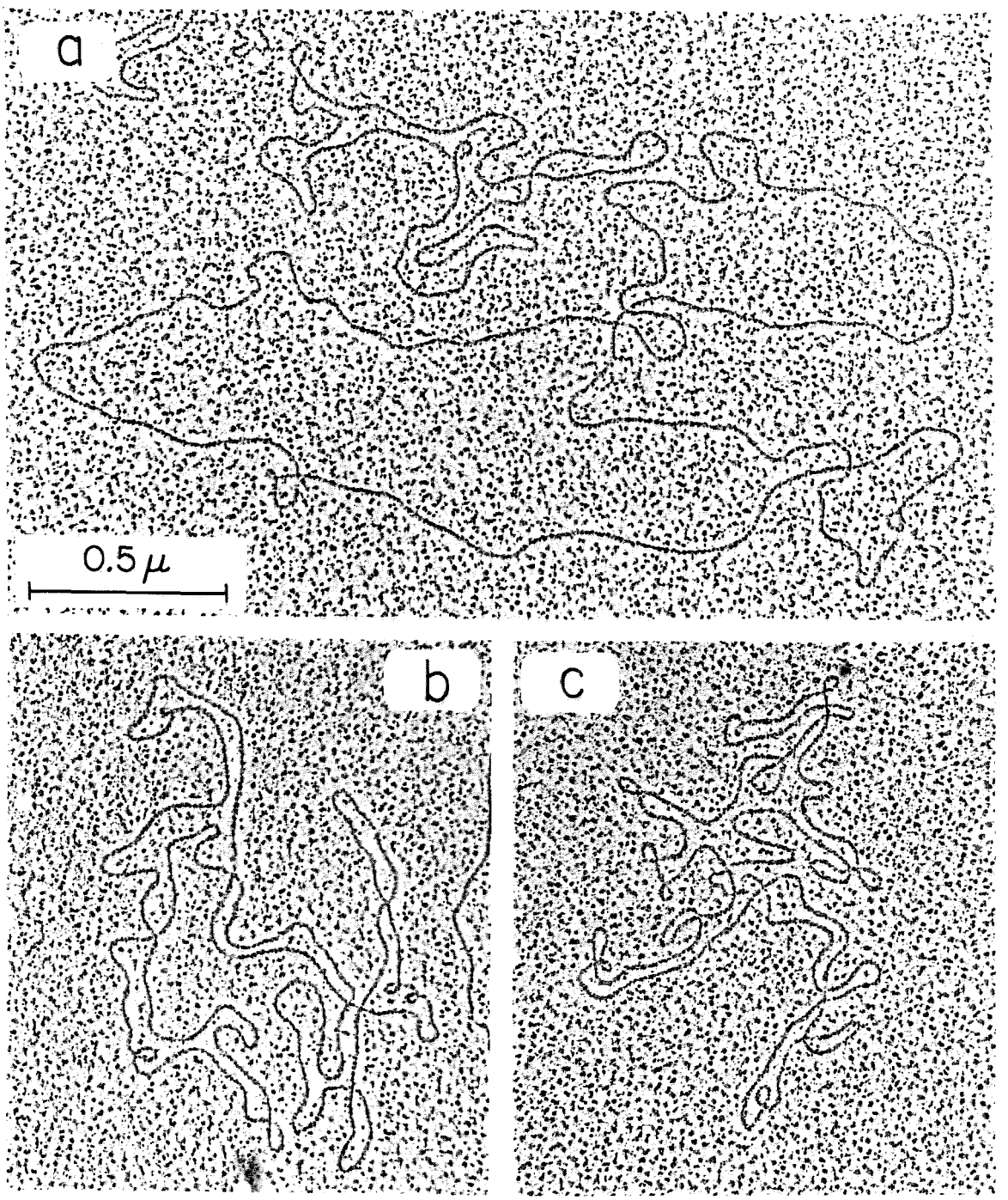

a
0.5 μ
b
c

Plate IV. Incomplete duplex molecules.

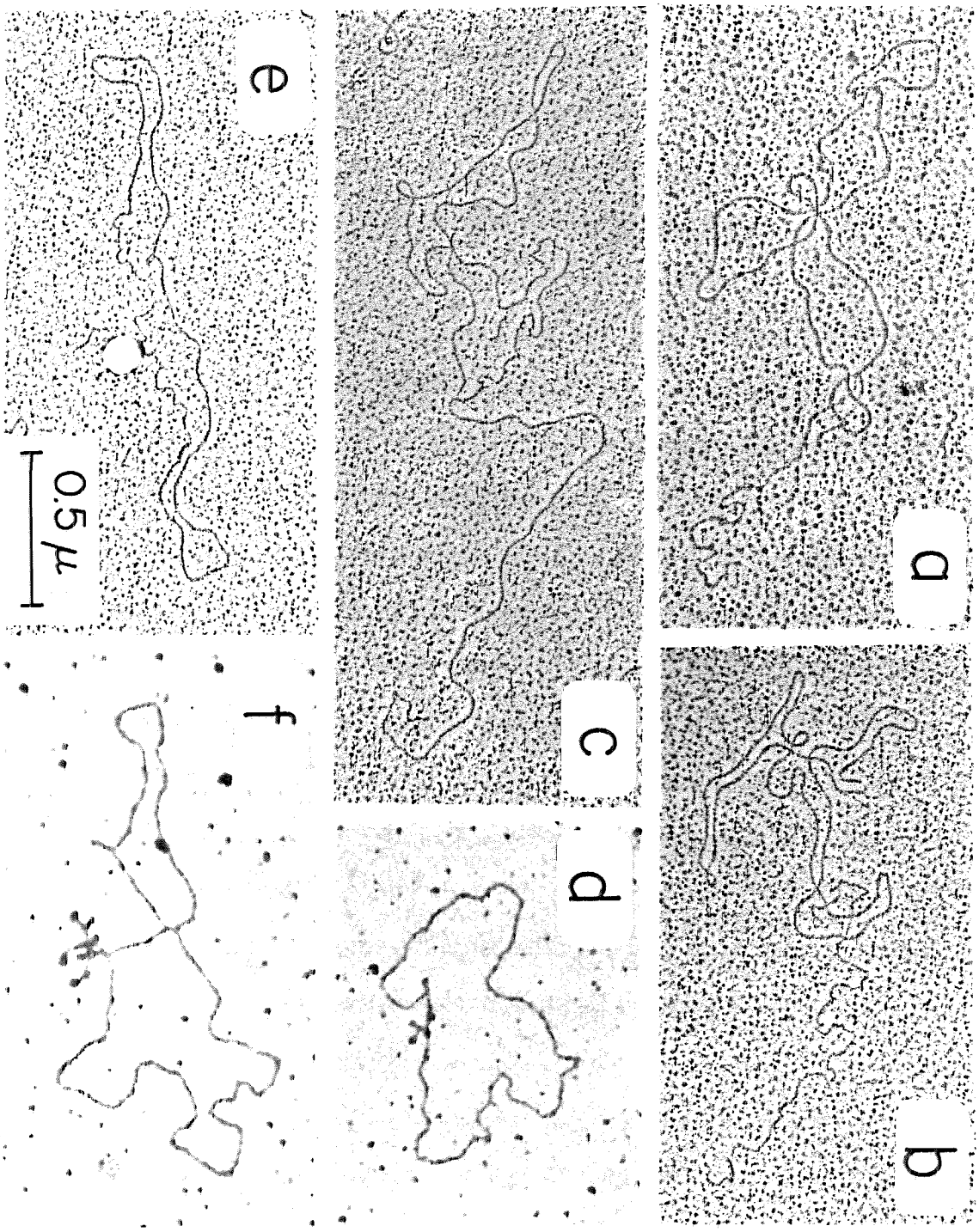
(a) and (b) Renatured circular duplex monomer molecules with a single-stranded tail. These molecules presumably correspond to the form, ($b_2$), in Figure 8, and, therefore, are intermediates in the formation of a fused dimer. The molecules were mounted on specimen grids by the formamide technique and shadowed with Pt-Pd.

(c) Renatured circular (partial) duplex monomer molecule with a duplex tail. The right end of the duplex tail appears to be single-stranded. The left end of the duplex tail appears to be attached to the circular monomer as a duplex; however, at the joint, the upper part of the circle appears to be single-stranded, while the lower part appears to be double-stranded. This molecule may be the result of renaturing a linear dimer with two monomers, one or more of which was a fragmented molecule. The DNA was mounted as in (a) and (b).

(d) Renatured molecule showing a single-stranded gap (Fig. 7(c)). This molecule is the result of renaturing a whole monomer strand with a short strand. The DNA was mounted by the aqueous technique and stained with uranyl acetate.

(e) Same as in (d), except mounted as in (a).

(f) Presumably the same type of intermediate that is shown in (a) and (b), but mounted as in (d).


a
b
c
d
e
f
0.5 μ

duplexes, can continue to undergo renaturation with the formation of branched molecules and large aggregates. The presence of such molecules considerably complicates the electron microscopic analysis. It is essential in these experiments, therefore, to utilize very lightly nicked DNA (one single-strand scission per duplex). As expected, incomplete duplex and aggregated molecules were observed in the electron microscope (Plates IV(d) and (e)). Since they were present in low relative frequency, they could be unambiguously identified, and did not interfere with the analysis of the renatured molecules derived from full-length single strands.

Another interesting class of incomplete duplexes are the intermediates in the formation of the heteroduplex from one dimer and one monomer strand. There were very few of these intermediates present, since the renaturation was carried almost to completion. A few examples that were found are shown in Plates IV(a), (b), and (f). These intermediates are composed of one complete duplex monomer circle with a single-stranded DNA tail and, presumably, correspond to the structure in Figure 8($b_2$). Also, a few monomer circular molecules with duplex tails were observed (Plate IV(c)).

The "figure 8" molecules were not formed when a sample of a nicked, but otherwise homogeneous, circular duplex DNA from PM2 virus (Espejo & Canelo, 1968) was reannealed. The nicked viral DNA (prepared as described in Materials and Methods, section (e)) was analyzed in alkaline CsCl by the band velocity procedure in the analytical ultracentrifuge. The DNA contained 20% circular strands and

50% homogeneously sedimenting linear strands. The remaining 30% was present as heterogeneous slowly moving linear strands. The nicked viral DNA with a molecular weight of about 6 million daltons was denatured and reannealed by the procedures used in the study of the monomer-dimer mitochondrial DNA system. The concentration during hybridization was 5 $\mu$g/ml. A second experiment was performed at 25 $\mu$g/ml. The first renatured sample contained 56% duplex circles, 15% PM2-size linear molecules, 16% large aggregates, 9% circles with tails, and 3% single-stranded molecules. The more concentrated renatured product consisted largely of aggregates. The remainder was a mixture of PM2-sized circular and linear duplexes and circles with tails. "Figure 8" molecules were not observed in the experiments with PM2 DNA. This result supports the validity of our classification of such molecules as monomer-dimer hybrids.

## 4. Discussion

This research was undertaken with the object of determining the structural relationship and homology between the circular monomers and circular dimers that occur in mixture in mitochondrial DNA samples prepared from leukocytes from patients with myelogenous leukemia. The study was carried out with M DNA from blood samples taken periodically from a single patient who had not been treated with cytotoxic drugs. The problems of homology and structure were

examined with two experimental approaches: (1) a centrifugal examination of the composition of the purified monomer and dimer species; (2) an examination of hybrids between monomers and dimers by electron microscope and centrifugal methods. Dawid & Wolstenholme (1968) employed a simpler version of the buoyant density procedure to detect qualitatively a sequence homology between Xenopus laevis and chick M DNA's. They also examined the renaturation products of purified X. laevis M DNA by electron microscope and centrifugal methods. They observed that the renatured products contained duplex circles with the contour length of native X. laevis M DNA, shortened duplex circles with single-strand regions, linear molecules of varying sizes, and aggregates.

### (a) Base composition

We have deduced from the buoyant behavior of separated monomers and dimers that these duplex species have the same base compositions. The equivalence of the buoyant densities in the analytical ultracentrifuge follows from the equivalence of the difference in buoyant density between each species and a marker DNA. The measurement error in each experiment corresponds to an error of approximately $\pm 0.5\%$ in the guanine-cytosine (GC) content of the DNA. The limit of detection for a difference is, therefore, $\pm 1\%$ in GC content.

The buoyant densities of the light and heavy strands in alkaline CsCl containing a marker DNA were also the same within the

accuracy of our measurement. In this case, we cannot clearly assess the quantitative aspects of the degree of similarity of base composition of the corresponding complements. The buoyant density difference (0.041 g/ml.) between light and heavy complements is due in part to a thymine bias between the strands. The dense strand (R. L. Hallberg, private communication) contains approximately 40% more thymidylate residues than does the light strand. Since all four types of residues contribute to the buoyant density of a titrated strand with unknown weighting factors, the experimental results—identical alkaline buoyant densities of light and heavy strands—could have been obtained by combinations of several differences in base composition that somehow compensate. On the other hand, the simpler interpretation strongly supported by the hybridization experiments is that the base compositions of the corresponding separate strands are, indeed, very similar.

(b) Homology

The results of experiments in which mixtures of light monomer and heavy dimer strands were reannealed and the buoyant density of the hybrid examined in the analytical ultracentrifuge clearly demonstrate that within the limits of detection these strands reanneal to form perfect hybrid duplexes. If the duplex had contained 10% single-stranded DNA, we should have observed a detectable shift of 0.001 g/ml. toward higher density. The results of these experiments are especially significant for the interpretation of structure of the various

molecular forms of hybridized DNA observed in the electron microscope.

The isolation of the separate complements from alkaline buoyant gradients centrifuged in the preparative ultracentrifuge induces a few scissions into each strand and provides materials suitable for the formation of the highly concatenated hybrid duplexes discussed above. The presence of such concatenated structures, which may be branched, interferes with the detectability of heterologous regions in the electron microscope. The electron microscope hybridization studies were, therefore, performed with mixtures of monomer and dimer molecules that had been carefully nicked and fractionated. Upon denaturation, this material then formed circular single strands and largely full-length linear strands. These mixtures were then reannealed and formed circular monomers, circular dimers, and fused dimers. Only the fused dimers are necessarily hybrid duplexes. These forms were separately inspected for the presence of heterologous regions and measured for contour length. A significant number of such molecules had the correct circular contour length and contained no heterologous regions. The absence of heterologous regions in these selected hybrid molecules reveals that deletions, additions, or insertions do not occur over lengths of DNA larger than 50 to 100 nucleotides. Smaller stretches, which do not give detectable heterologous regions, could, of course, occur.

If we now consider both the electron microscope and centrifuge

analysis of the hybrids together, we conclude that the total heterology cannot exceed 10%; and, if any heterology at all is present, it must be dispersed in short noncontiguous sections of less than 50 to 100 nucleotides.

(c) The head-to-tail structure of the circular dimer

The strongest evidence for the head-to-tail arrangement of monomer genome in the circular dimer was obtained in the electron microscope experiment with the slowly quenched denatured DNA sample. If the dimer contained a head-to-head arrangement, the linear dimer strands should have, by this treatment, annealed to 5-$\mu$ linear duplexes with hairpins at both ends. Since no linear duplex molecules were observed in this experiment in which a small number of circular monomer duplexes formed, we conclude that the monomer sequences are arranged in the form of a head-to-tail circular concatenate. Five-micron linear duplexes were also not observed in the extensive study of the completely reannealed DNA.

(d) The homogeneity of base sequence in monomeric mitochondrial DNA

Mitochondrial DNA in animals, from the sea urchin to man, occurs in the form of circular duplex molecules approximately 5 $\mu$ in contour length and in the form of multiples of this unit length. The unit length contains about 16,000 nucleotide pairs. This DNA, if homogeneous and if engaged entirely in the formation of mRNA

for protein synthesis, can specify no more than 5,300 amino acid residues, a number that is inadequate to account for the synthesis of a complete mitochondrion. Most cells, however, contain a large number of M DNA molecules (e.g., 250 to 500 in a HeLa cell) and each could conceivably be different, a possibility that is unlikely in view of the well established rapid renaturability of M DNA. The standard method at the present time for determining the amount of genetic information in a unit mass of DNA is based on a correlation between the second-order rate constant for annealing of denatured DNA under standard conditions and the complexity of the DNA (Britten & Kohne, 1965; Wetmur & Davidson, 1968). Borst, Ruttenberg & Kroon (1967) performed such measurements and reported that the apparent genome size of chick liver M DNA was approximately eight-tenths the size of the molecule. This result indicates that chick liver M DNA is essentially homogeneous.

The results described in this study, while focused on the monomer-dimer hybrids, also yield significant information regarding the homogeneity of the leukocyte M DNA monomer and dimer species individually. The complete reannealing of the denatured M DNA mixture, as observed in the buoyant density experiment (Fig. 7(c)), and the failure to observe heterologous regions in the renatured monomers or the renatured dimers (Plates I through IV), clearly demonstrates that the monomer and dimer species are not in themselves mixtures of partially homologous DNA's. Either these DNA's are essentially homogeneous or they consist of classes of DNA's that

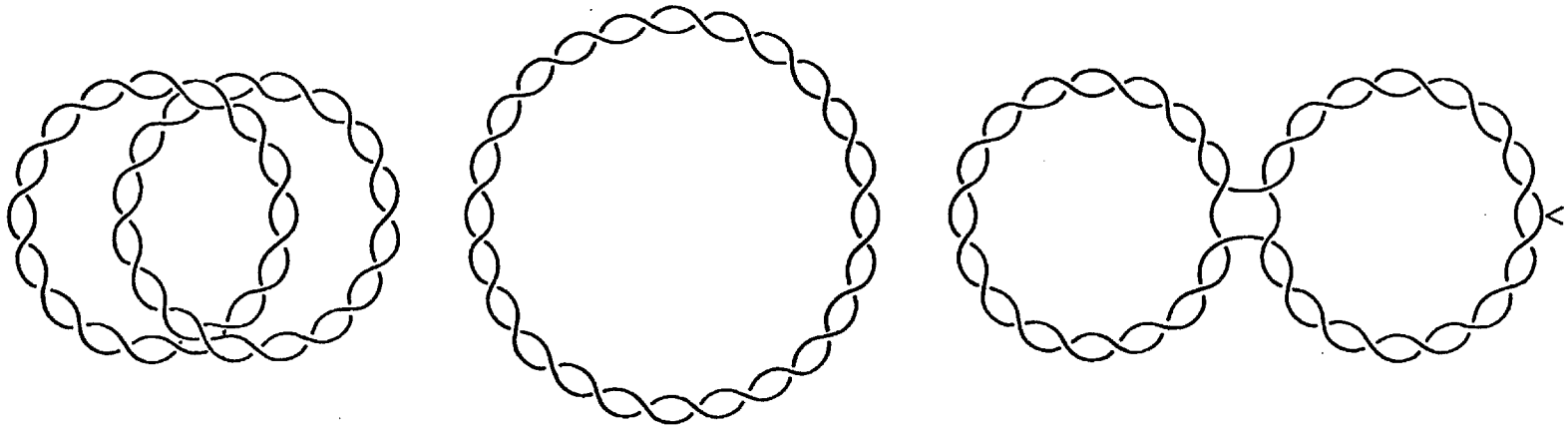
are so slightly homologous that they do not reanneal with each other. The latter alternative is held to be less likely in view of the further requirements that the DNA's in the putative classes also must have the same single-strand base composition and molecular weight. Both the above results and the kinetic result of Borst, Ruttenberg & Kroon (1967) suggest that M DNA should be regarded as homogeneous with an information content that corresponds to its molecular weight. The presence of point mutations or small deletions or insertions that could not have been detected are, of course, not ruled out.

(e) Three structural forms of dimeric mitochondrial DNA

Figure 9 presents three different structural forms of dimeric M DNA diagrammatically. The topologically bonded catenated dimer (Fig. 9(a)) and the circular dimer (Fig. 9(b)) occur naturally. The fused dimer prepared in small amounts in this work contains a four-stranded junction region represented in Figure 9(c) with five unpaired bases in each strand. The actual number of unpaired bases in the junction regions is not known at present. An inspection of two juxtaposed space-filling duplex models suggested that the junction could be formed with less than five unpaired bases per strand. The junction region in the fused dimer formally corresponds to a fourfold or X-type branch from which four duplexes emanate. It may be contrasted with the threefold or Y-type branch point postulated as a formal intermediate

Figure 9. Representation of three structural isomers of dimeric mitochondrial DNA. The fused dimer is shown as a duplex containing a linear dimer strand and two circular monomer strands. The carat (v) represents a single-strand scission.

STRUCTURAL ISOMERS OF DIMERIC
MITOCHONDRIAL DNA



CATENATED
DIMER

CIRCULAR
DIMER

FUSED
DIMER



Figure 9

at the origin and, also, at the growing point of DNA partially replicated by a semiconservative mechanism as discussed by Cairns (1963). The completion of replication of a circular duplex by the Cairns mechanism formally involves a fusion of the origin and the growing point and the formation of a fourfold junction region.

Bloomfield (1966) and Fukatsu & Kurata (1966) have calculated the effect of permanently associating short regions in a circular duplex in such a way as to form molecules which contain equal-sized duplex loops. These calculations were performed to assess the hydrodynamic effects of superhelical turns in a closed circular duplex. They calculated that a molecule containing two loops should sediment about 15 percent faster than the corresponding parent molecule, a result which can be checked experimentally when larger quantities of fused dimers are prepared and purified.

(f) "Normal" and "abnormal" mitochondrial DNA

As stated in the introduction, we have referred to the circular dimer as an "abnormal" form, because it occurred in leukemic leukocytes and did not occur in the appropriate controls. It must be emphasized that the present study compares the information content of the circular dimer to the circular monomers in the same patient. A similar study of the homology relationships between circular dimers from a leukemic patient and circular monomers from a nonleukemic (i. e., "normal") person remains to be performed.

## Summary

Mitochondrial DNA occurs in multiple circular forms in human leukemic leukocytes. As part of an inquiry into the structural relationships between the forms, the 5- and 10-$\mu$ circular duplexes were separated and characterized by sedimentation procedures. The buoyant densities in neutral cesium chloride were indistinguishable, indicating that overall base compositions of the two species are the same. The buoyant densities of the corresponding heavy and light strands in the two species were also indistinguishable. Together, the results suggest that the dimer consists of two monomer genomes.

Reannealing of equal amounts of fragmented light monomer and heavy dimer complementary strands isolated from alkaline buoyant cesium chloride gradients leads to a high molecular weight duplex with a buoyant density indistinguishable from the parent duplexes. It is concluded from this result that the degree of homology between monomer and dimer is at least 90%.

Denaturation of a mixture of lightly nicked monomer and dimer duplexes, followed by reannealing in formamide solutions, leads to mixtures of circular monomer and dimer duplexes and heteroduplexes which contain one dimer strand and two monomer strands. These heteroduplexes occur in the form of a "figure 8" molecule, a new DNA structure which we have called the fused dimer. Neither heterologous regions, nor insertions, nor deletions were found in a careful examination of the electron micrographs.

It is concluded from this part of the work that heterologous regions, insertions, or deletions exceeding 50 to 100 nucleotides in length do not occur in the heteroduplexes. At this level of insight, the dimer is a circular concatamer of two monomer genomes. In addition, the results show that leukemic leukocyte mitochondrial DNA monomers (as well as dimers) are substantially homogeneous in base sequence.

The electron microscope studies also show that monomer genomes in the dimer are connected in a head-to-tail structure, rather than in a head-to-head structure.

## ACKNOWLEDGEMENTS

We wish to thank R. L. Hallberg for helpful discussions, for providing us with separated HeLa M DNA strands, and for permission to publish his results in Figure 5(a). We also thank M. Teplitz for her help in the isolation of M DNA. This work was supported in part by grants no. CA08014, GM15327, GM01262, and GM06965 from the National Institutes of Health, United States Public Health Service.

This is contribution number 3894 from the Division of Chemistry and Chemical Engineering, California Institute of Technology.

## REFERENCES

Bauer, W. & Vinograd, J. (1968). J. Mol. Biol. 33, 141.

Bloomfield, V. A. (1966). Proc. Nat. Acad. Sci., Wash. 55, 717.

Borst, P., Ruttenberg, G. J. C. M. & Kroon, A. M. (1967). Biochim. biophys. Acta. 149, 140.

Britten, R. J. & Kohne, D. E. (1966). Yearb. Carnegie Inst. 1965, p. 78.

Cairns, J. (1963). J. Mol. Biol. 6, 208.

Clayton, D. A., Smith, C. A., Jordan, J. M., Teplitz, M. & Vinograd, J. (1968). Nature, 220, 976.

Clayton, D. A. & Vinograd, J. (1967). Nature, 216, 652.

Clayton, D. A. & Vinograd, J. (1969). Proc. Nat. Acad. Sci., Wash. 62, 1077.

Corneo, G., Zardi, L. & Polli, E. (1968). J. Mol. Biol. 36, 419.

Davidson, N., Widholm, J. M., Nandi, U. S., Jensen, R. H., Olivera, B. M. & Wang, J. C. (1965). Proc. Nat. Acad. Sci., Wash. 53, 111.

Davis, R. W. & Davidson, N. (1968). Proc. Nat. Acad. Sci., Wash. 60, 243.

Davis, R. W., Simon, M. & Davidson, N. (1969). Manuscript in preparation.

Dawid, I. B. & Wolstenholme, D. R. (1968). Biophys. J. 8, 65.

Espejo, R. T. & Canelo, E. S. (1968). Virology, 34, 738.

Fukatsu, M. & Kurata, M. (1966). J. Chem. Phys. 44, 4539.

Hershey, A. D. & Burgi, E. (1965). Proc. Nat. Acad. Sci., Wash. 53, 325.

Hudson, B., Clayton, D. A. & Vinograd, J. (1969). Cold Spr. Harb. Symp. Quant. Biol. 33, 435.

Rownd, R., Lanyi, J. & Doty, P. (1961). Biochim. biophys. Acta. 53, 225.

Schildkraut, C. L., Marmur, J. & Doty, P. (1962). J. Mol. Biol. 4, 430.

Thomas, C. A., Jr. & MacHattie, L. A. (1964). Proc. Nat. Acad. Sci., Wash. 52, 1297.

Vinograd, J. & Hearst, J. E. (1962). In Fortschritte de Chemie Organischer Naturstoffe, ed. by L. Zechmeister, vol. 20, p. 372. Vienna: Springer-Verlag.

Vinograd, J., Morris, J., Davidson, N. & Dove, W. F., Jr. (1963). Proc. Nat. Acad. Sci., Wash. 49, 12.

Watson, R., Bauer, W. & Vinograd, J. (1969). Biophys. J., Society Abstracts, 9, A-225.

Westmoreland, B. C., Szybalski, W. & Ris, H. (1969). Science, 163, 1343.

Wetmur, J. G. & Davidson, N. (1968). J. Mol. Biol. 31, 349.

# CHAPTER 7

## Base Sequence Heterogeneity in SV40 DNA

## 1. Introduction

It has been recently reported that successive passage of SV40 virus at high multiplicity in cell cultures results in the production of heterogeneous defective particles of lower densities than the infectious virus (Yoshiike, 1968a; Yoshiike, 1968b; Uchida & Watanabi, 1968). These buoyant density mutants have altered plaque-forming (PF) ability, and altered T antigen-forming (TA) and V antigen-forming (VA) activity. The defect appears to be contained in the DNA since the activity (PF, TA, & VA) of the DNA extracted from the buoyant density mutants parallels the activity of the mutant virus. The defective SV40 viral DNA has also been shown by velocity sedimentation and electron microscopy to be somewhat shorter than the infectious viral DNA (Yoshiike, 1968a; Yoshiike, 1968b; Uchida et al, 1968). Similar results have been obtained with polyoma virus DNA (Thorne, 1968; Blackstein, Stanners & Farmilo, 1969). It has therefore been suggested that these mutants of SV40 and of polyoma are the result of a deletion(s) in their DNA and are possibly similar to $\lambda$ deletion mutants.

Some of these SV40 mutants have greatly increased transforming activity (Uchida et al, 1968; Todaro & Takemoto, 1969). Therefore, a study of the type, extent, and location of the DNA base sequence changes in SV40

DNA will greatly aid in the understanding of the genetic material governing transformation. That base sequence changes in DNA can be easily studied by electron microscopy is the theme of this thesis.

## 2. Materials and Methods

### (a) SV40 DNA

The SV40 DNA was obtained from the laboratory of J. Vinograd and had been grown by serial passage at high multiplicity of infection.

### (b) Nicking of closed circular SV40 DNA

As shown in the study of the renaturation of circular mitochondrial DNA (Clayton, Davis & Vinograd, 1969, and included in this thesis), it is necessary to introduce one, and, if possible, no more than one, single-strand scission per double-stranded circular SV40 DNA. Lightly nicked SV40 DNA was obtained by very mild DNase treatment (a gift from W. Upholt) or was obtained as an upper band from a CsCl-EB density gradient of a stock of SV40 DNA (a gift from H. Gray).

### (c) Hybridization

Hybridizations of the lightly nicked SV40 DNA were performed as described by Clayton et al (1969) or as described by Parkinson & Davis (1969). Both methods gave the same results.

### (d) Electron microscopy

The DNA was mounted for electron microscopy as described in Davis, Simon & Davidson (1969) and in Clayton et al (1969). A hyperphase containing 0.5 $\gamma$/ml of DNA, 0.1 mg/ml cytochrome c in 0.1 M ammonium acetate, 0.01 M

tris (pH 8) and 30% or 60% formamide was layered onto a hypophase of 0.01 M tris and 10% or 40% formamide respectively. The 60% onto 40% formamide combination was used to increase the visibility of the small non-homology regions. These conditions do not cause the partial denaturation of duplex λDNA and therefore only true non-homology regions should be opened.

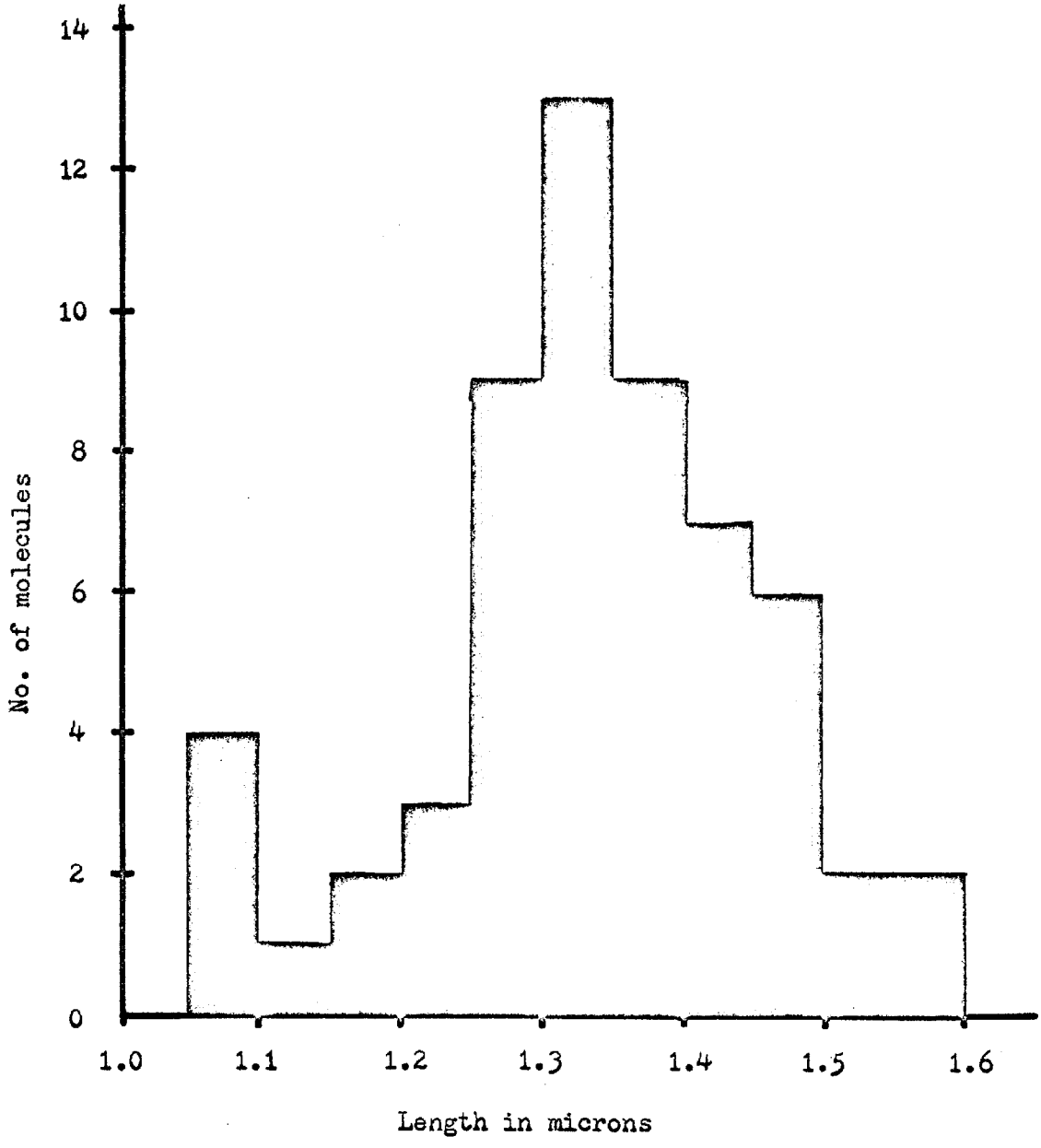
## 3. Results and Discussion

### (a) Length heterogeneity of SV40 DNA

The molecular length heterogeneity of this sample of SV40 DNA was determined by electron microscopy as described in Davis, Simon, & Davidson (1969) and in Lee, Davis & Davidson (1969). The distribution of molecular lengths is shown in Fig. 7-1. The distribution contains only one major peak at 1.34 $\mu$ and does not appear to be skewed toward shorter lengths as might have been expected if the SV40 DNA sample contained a small amount of short defective viral DNAs. However, the percent standard deviation of the distribution is 8.8% and is in rough agreement with the results of Yoshiike (1969a). The expected percent standard deviation for a homogeneous DNA population of the length of SV40 DNA is $3.7//\overline{0.84} = 4.0\%$ (Davis et al, 1969). Therefore, the variability, expressed as a percent standard deviation, in the number of base pairs in SV40 DNA is $(8.8^2 - 4.0^2)^{\frac{1}{2}} = 7.9\%$.

If there is a homogeneous DNA sample contained in the distribution of Fig. 7-1, 95% of these molecules should be contained within $\pm$ 0.1 $\mu$. Therefore, if the heterogeneity in SV40 DNA is due solely to DNA deletion(s), then the molecules shorter than about 1.4 $\mu$ should contain a DNA deletion(s) or 70% of the DNA molecules in this sample of SV40 should contain a DNA deletion(s). However,

Fig. 7-1. Length distribution of SV40 DNA. The length distribution of SV40 DNA was determined by electron microscopy. Since no convenient DNA calibration marker was available, the length of the DNA was calibrated with a diffraction grading and expressed in microns. In order to prevent magnification errors, the molecules on a single, densely populated grid square were photographed without refocusing the electron microscope. The mean of the distribution is 1.34$\mu$ and the fractional standard deviation of the distribution is 8.8%. The expected fractional standard deviation for a homogeneous DNA population of this length is 4.0%.


14
12
10
8
6
4
2
0
No. of molecules
1.0
1.1
1.2
1.3
1.4
1.5
1.6
Length in microns

the heterogeneity in SV40 DNA could also be due to DNA additions, to the occurance of several non-homologous DNA molecules of different lengths, to DNA substitutions in which a portion(s) of the SV40 genome has been replaced by a different amount of foreign DNA, or to a combination of these possibilities. An electron microscopic study of self-hybridized SV40 DNA can distinguish between these several possibilities.

(b) Base sequence changes in SV40 DNA

In order to determine the nature of the DNA base sequence changes in SV40 DNA a sample of lightly nicked SV40 DNA was denatured by either 95% formamide or 0.1 M NaOH and renatured in 50% formamide solution at pH 8. The renatured DNA was mounted for electron microscopy by both the aqueous and the formamide techniques. Plates 7-I and 7-II show examples of some of the more interesting renatured molecules. Quite surprisingly, there is a wide diversity of renatured molecular forms containing non-homology regions. The diversity was so great that it was impossible to define unique classes of molecules. However, it can be said that at least 50% of the renatured molecules contain non-homology regions. A few loops of single-stranded DNA were seen that are characteristic of deletions (Plate 7-Ib) as has been predicted (Yoshiike, 1968a, Yoshiike; 1968b; Blackstein et al, 1969). However, most of the renatured molecules contained loops character-

Plate 7-I Renatured SV40 DNA

(a) The renatured DNA was mounted by the aqueous technique. Two regions of single-stranded DNA (bushes) are visible. The DNA was stained with uranyl acetate.

(b) The renatured DNA was mounted by the formamide technique (60% onto 40% formamide). The non-homology region appears to be of the deletion or addition type. The loop of single-stranded DNA does not originate from a single point on the double-stranded circle. However, in the high formamide concentration that was used, deletion loops in heteroduplexes often have this appearance. The DNA was stained with uranyl acetate.

(c) & (d) The renatured DNA was mounted as given in (b). The non-homology region appears to be a single large substitution. The size of the non-homology region corresponds to one or two genes. The DNA was shadowed with platinum-palladium.

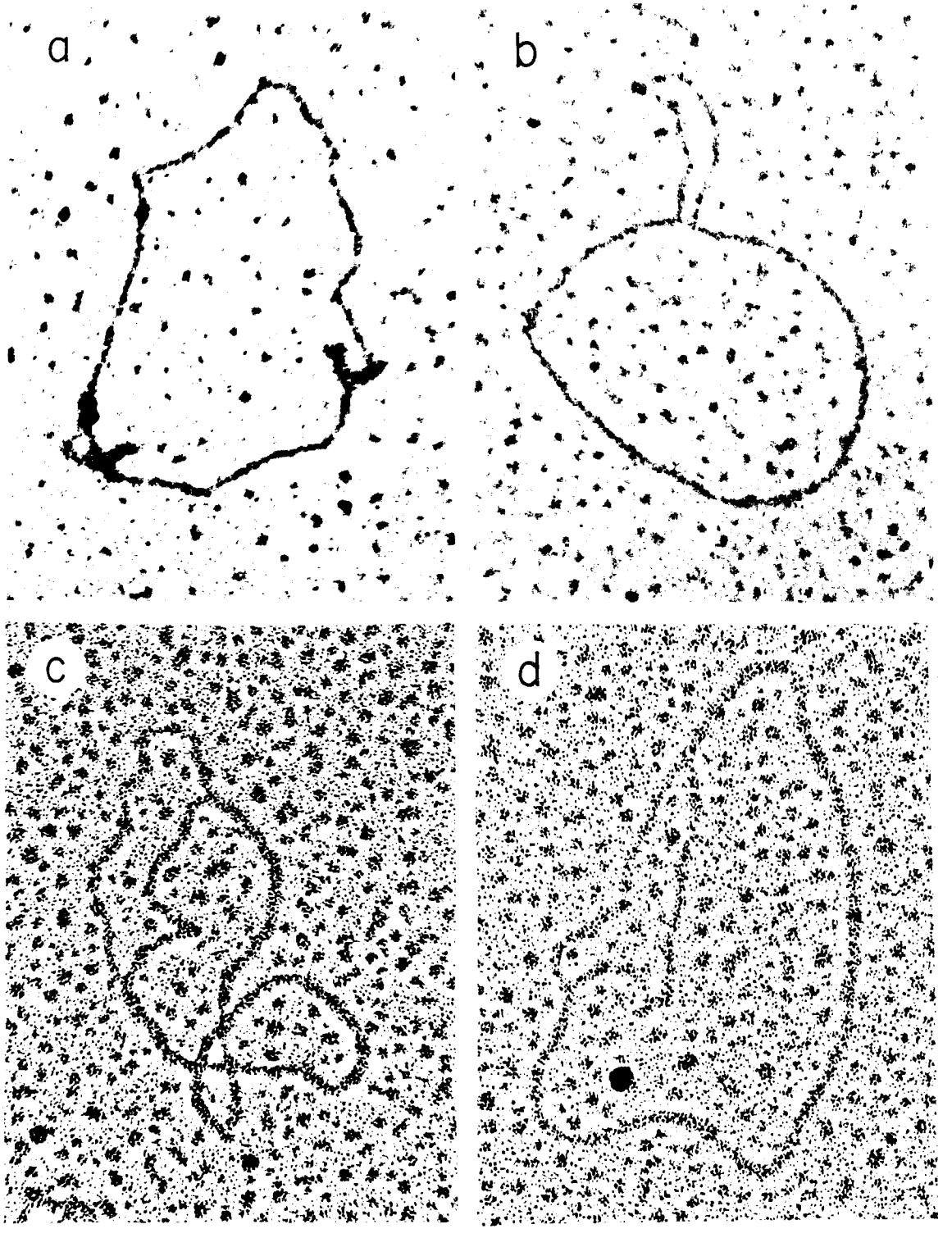

a
b
c
d

Plate 7-II Renatured SV40 DNA with multiple substitutions

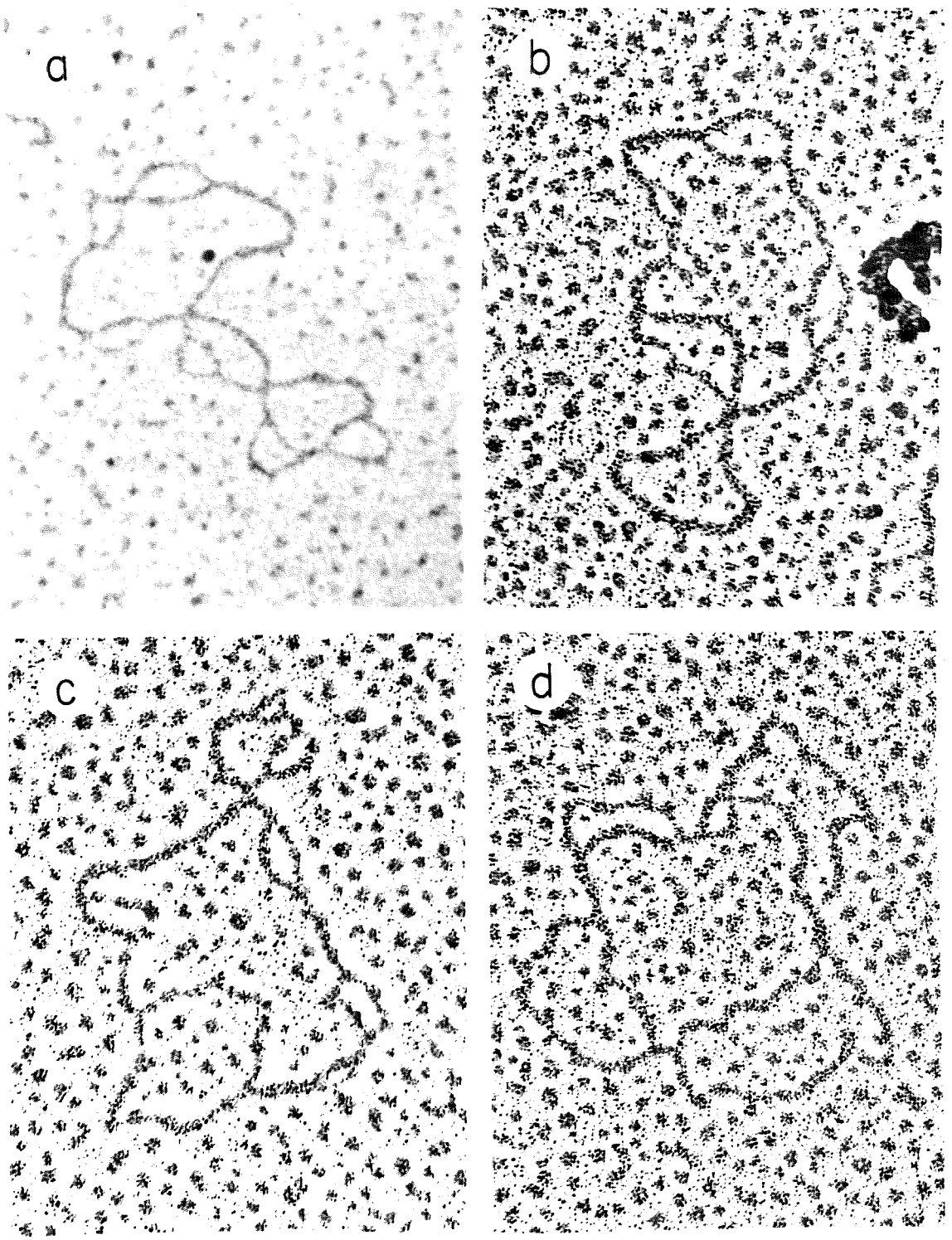
(a) The renatured DNA shows at least four small substitution type non-homology regions. Note that the two single-strands of DNA in each non-homology region are about equal in length and are each equal to or are slightly less than one gene.

(b) The same as (a) except the relative locations of the non-homology regions appear to be different.

(c) The same as (b) except there are at least five non-homology regions.

(d) Either "renatured" DNA that is very non-homologous, or the hybridization of two complementary single-stranded circular molecules. Note that the loops are about of equal size. Since lateral aggregation is possible in the high formamide concentration that was used to mount the DNA, the appearance of this molecule may result from the overlap and slight lateral aggregation of two single-stranded circular molecules.

The DNA was mounted by the formamide technique (60% onto 40% formamide) and stained with uranyl acetate. Plates (b), (c), & (d) were also shadowed with platinum-palladium.


a
b
c
d

istic of substitutions (Plate 7-Ic & d and Plate 7-II).

A fraction of the renatured DNA population appeared to contain a single large substitution or a single deletion (addition). Three examples are shown in Plate 7-Ib, c, & d.

Another fraction of the renatured DNA population contained multiple substitution loops, most of them from 150 to 500 base pairs in length (Plate 7-II). The number of substitution loops per renatured molecule was variable but often contained 4 or 5.

Because of the wide diversity of renatured forms, it is impossible to draw any conclusions as to the origin of the non-homologous DNA in this sample of SV40 DNA. Since unique classes of molecules could not be clearly defined, the conclusions drawn from a micrograph of any one molecule are dubious.

The base sequence changes in SV40 DNA could have occurred through the incorporation of host DNA into the viral genome. It should be noted, however, that non-homology regions characteristic of substitutions (Plate 7-Ic and d) could arise from the hybridization of two partially overlapping or adjoining deletions. For example, the double deletion heteroduplex ($\lambda b_2/\lambda b_{508}$) appears as a substitution loop. The renatured molecules which contained multiple non-homology regions could be the result of the incorporation of small multiple substitutions or

the incorporation of a single large, partially sequence homologous, substitution of host DNA. These types of renatured molecules also appear to be somewhat similar to heteroduplexes between different lambdoid phages (Simon & Davidson, 1969). Therefore, there might be several different but closely related viruses in the SV40 sample. However, the non-homology regions in lambdoid phage heteroduplexes are usually at least gene size and much larger than the non-homology regions seen in renatured SV40 DNA.

An interesting observation is that the renatured molecules with multiple non-homology regions were invariably circular with no free single-stranded or double-stranded ends. This fact is quite puzzling since all renatured molecules are presumably formed either from two linear strands or from one linear and one circular strand. Since no free ends were observed, the ends of the linear strands are always within regions of homology. This seems quite unlikely. Plate 7-IId shows a molecule with very little duplex DNA but with no free ends. The multiple non-homology site hybrid molecules are also usually tangled with many super-twists, somewhat analogous to super-coiled closed circles. It is therefore speculated that these molecules are the result of the renaturation of two circular single strands.

It has been assumed that two complementary single-stranded circular molecules could not hybridize. However,

if hybridization were to occur, for every Watson-Crick right-handed turn of the helix, there must also be an accompanying left-handed turn somewhere in the circle. The stability and exact nature of the left-handed helix is not clear. In this case, the left-handed helix might gain stability from the constraints imposed by the stable Watson-Crick helix. Under the mild denaturing conditions used here for mounting the DNA for electron microscopy, the right/left helix hybrids might be partially stable. The possibility of hybrids between complimentary circular molecules could be eliminated if the DNA sample is lightly nicked with DNase after hybridization. Circular/circular DNA hybrids would instantaneously renature once a single-strand break is introduced into one of the circles.

In order to carry this study further it is necessary to obtain a homogeneous sample of SV40 DNA. The introduction of non-homogeneous DNA into the SV40 genome can then be followed by studying samples after each serial passage.

## References

Blackstein, M. E., C. P. Stanners and A. J. Farmilo (1969). J. Mol. Biol. 42, 301.

Clayton, D. A., R. W. Davis and J. Vinograd (1969). To be published in J. Mol. Biol. and is included in this thesis.

Davis, R. W., M. Simon and N. Davidson (1969). To be published in Methods in Enzymology and included in this thesis.

Lee, C. S., R. W. Davis and N. Davidson (1969). To be published in J. Mol. Biol. and is included in this thesis.

Parkinson, J. S. and R. W. Davis (1969). To be published in J. Mol. Biol. and is included in this thesis.

Simon, M., and N. Davidson (1969). To be published.

Thorne, H. V. (1968). J. Mol. Biol. 35, 215.

Todaro, J. and K. Takemato (1969). Proc. Nat. Acad. Sci., Wash. 62, 1031.

Uchida, S. and S. Watanabe (1968). Virology, 35, 166.

Yoshiike, K. (1968a). Virology, 34, 391.

Yoshiike, K. (1968b). Virology, 34, 402.

# CHAPTER 8

## Single-stranded ØX174 DNA Synthesis: Electron Microscopy of the Replicating Intermediates

This work was done in conjunction with Rolf Knippers and composes a part of the paper "The Process of Infection with Bacteriophage ØX174 XXIX. *In Vivo* Studies on the Synthesis of the Single-stranded DNA of Progeny ØX174 Bacteriophage", Knippers, Razin, Davis, & Sinsheimer (1969). *Journal of Molecular Biology*, *45*, 237. The results and discussion given here are largely excerpted from this paper.

## Summary

Double-stranded ØX DNA rings (RF), composed of a closed complementary strand and an open viral strand, are found in ØX-infected cells during the period of progeny single-stranded DNA formation. The linear viral strand components of these double-stranded circles are "nicked" between a deoxyguanosine (on the 3'-terminus) and a deoxycytidine (on the 5'-terminus). The 5'-terminus is not phosphorylated. Upon replication, nucleotides are added to the 3'-end of the viral strand. The 5'-end is concomitantly displaced from its complementary strand template and appears, by electron microscopy and by column chromatography, as a single-stranded, non-hydrogen-bonded, tail of a double-stranded circle.

While the viral strand grows (to a maximum of twice the length of one ØX genome) by addition of nucleotides to the 3'-end, an interesting reaction, of uncertain nature, also takes place at the free 5'-end of the nascent viral strand, which has the net result that the original terminal deoxycytidine is replaced by a terminal deoxyguanosine.

Our experiments indicate that the single-stranded DNA synthesis does not take place at the cell membrane where double-stranded ØX DNA is formed but upon "cytoplasmic" RF. Viral coat protein can be shown to be associated with the single-strand-producing complexes.

## 1. Introduction

When cells of E. coli are infected with the bacteriophage ØX174, three stages of viral DNA synthesis have been distinguished: (i) the conversion of the infecting single-stranded DNA to a double-stranded form, the parental "replicating form" (RF) DNA of ØX* (Sinsheimer, 1961; Sinsheimer, Starman, Nagler & Guthrie, 1962; Hayashi, Hayashi, & Spiegelman, 1963; Matsubara, Taketo & Takagi, 1963); (ii) the semiconservative replication of the parental RF with the production of ten to twenty daughter RF molecules per infected cell (Yarus & Sinsheimer, 1967; Denhardt & Sinsheimer, 1965; Stone, 1967; (iii) the formation of progeny single-stranded DNA (Sinsheimer, Knippers & Komano, 1968). The RFII molecules* serve as precursors for the single-stranded progeny ØX DNA which is produced in a semiconservative asymmetric manner using as templates the complementary strands (of the daughter RFs) formed during the second stage (Lindquist & Sinsheimer 1967; Knippers, Komano & Sinsheimer, 1968; Komano, Knippers & Sinsheimer, 1968). The primary object of this study was an electron microscopic identification of the replicating intermediates in single-stranded DNA formation.

DNA was gently extracted from infected cells during single-stranded DNA replication by the sarkosyl-pronase procedure, and the replicating intermediates were separated from the larger host DNA by neutral sucrose sedimentation

(Knippers, Razin, Davis & Sinsheimer, 1969). The leading edge of the RFII band was examined in the electron microscope for the replicating intermediates. (As a control, various positions in the gradient were also examined.)

Footnote

*RF is the double-stranded "replicative form" DNA of ØX. In RFI (supercoiled) both strands are circular and closed. In RFII one or both strands of the circle are nicked. All the RFII molecules described in this section are composed of a closed circular complementary strand and a nicked viral strand.

## 2. Methods

Grids for electron microscopy were prepared by the basic protein film technique (Kleinschmidt & Zahn, 1959) and stained with uranyl salts or shadowed with platinum-palladium. Under aqueous conditions single-stranded DNA is collapsed and appears as a "bush", while double-stranded DNA appears as a linear filament (Davis & Davidson, 1968). However, when mounted under mild denaturing conditions (30% formamide), single-stranded DNA appears as a flexible thin linear filament, while double-stranded DNA and protein are generally unaffected. Both of these mounting techniques were used.

The aqueous mounting technique is essentially that described in Davis and Davidson (1968). 30 $\mu$l of a solution containing about 0.2 $\mu$g/ml of DNA and 0.1 mg/ml of cytochrome c, in 0.5 M ammonium acetate, pH 7, were spread onto 0.25 M ammonium acetate, pH 7. The grids were stained with uranyl acetate.

The denatured single-strand DNA mounting technique was developed from the excellent technique described by B. Westmoreland (personal communication). 50 $\mu$l of a solution containing about 0.2 $\mu$g/ml of DNA and 0.1 mg/ml of cytochrome c, in 0.1 M ammonium acetate, 0.01 M Tris, pH 8, and 30% formamide were spread onto 0.01 M ammonium acetate, $10^{-3}$ M Tris, pH 8, and 10% formamide. Both of these solutions were prepared one min before

spreading, and the film was picked up on specimen grids two min after spreading. Grids were stained with uranyl acetate and shadowed with platinum-palladium.

Micrographs were made with a Phillips EM 300 electron microscope using a 50 $\mu$ objective aperture and 60-KV accelerating voltage.

## 3. Results

Upon electron microscopic examination, two general classes of DNA were observed: ØX174 RF DNA and fragmented linear E. coli DNA. Although the ØX174 RF DNA comprised only 10 to 50% (by number) of the total DNA molecules, it was easily distinguished from the larger linear E. coli DNA.

When the DNA was mounted using the aqueous technique, approximately 20% of the RFII molecules had a single-stranded DNA "bush" attached to them (Plate 8-Ia). The bush appearance is derived from the fact that single-stranded DNA is collapsed on itself through random base interactions (Davis & Davidson, 1968). This type of base interaction is easily melted out with formamide. Therefore, to determine more clearly the nature and size of the single-stranded DNA branch on the RFII molecules, the replicating intermediates were mounted for electron microscopy in the presence of 30% formamide (at this concentration of formamide there is little effect on double-stranded DNA or protein). Upon examination of the replicating intermediates mounted under these mild denaturing conditions, about 20% of the double-stranded RFII molecules had a linear single-stranded DNA branch attached to them at only one position (Plate 8-I b,c,& d). The length of the single-stranded DNA branch was always found to be shorter than that of the mature single-stranded circular viral DNA

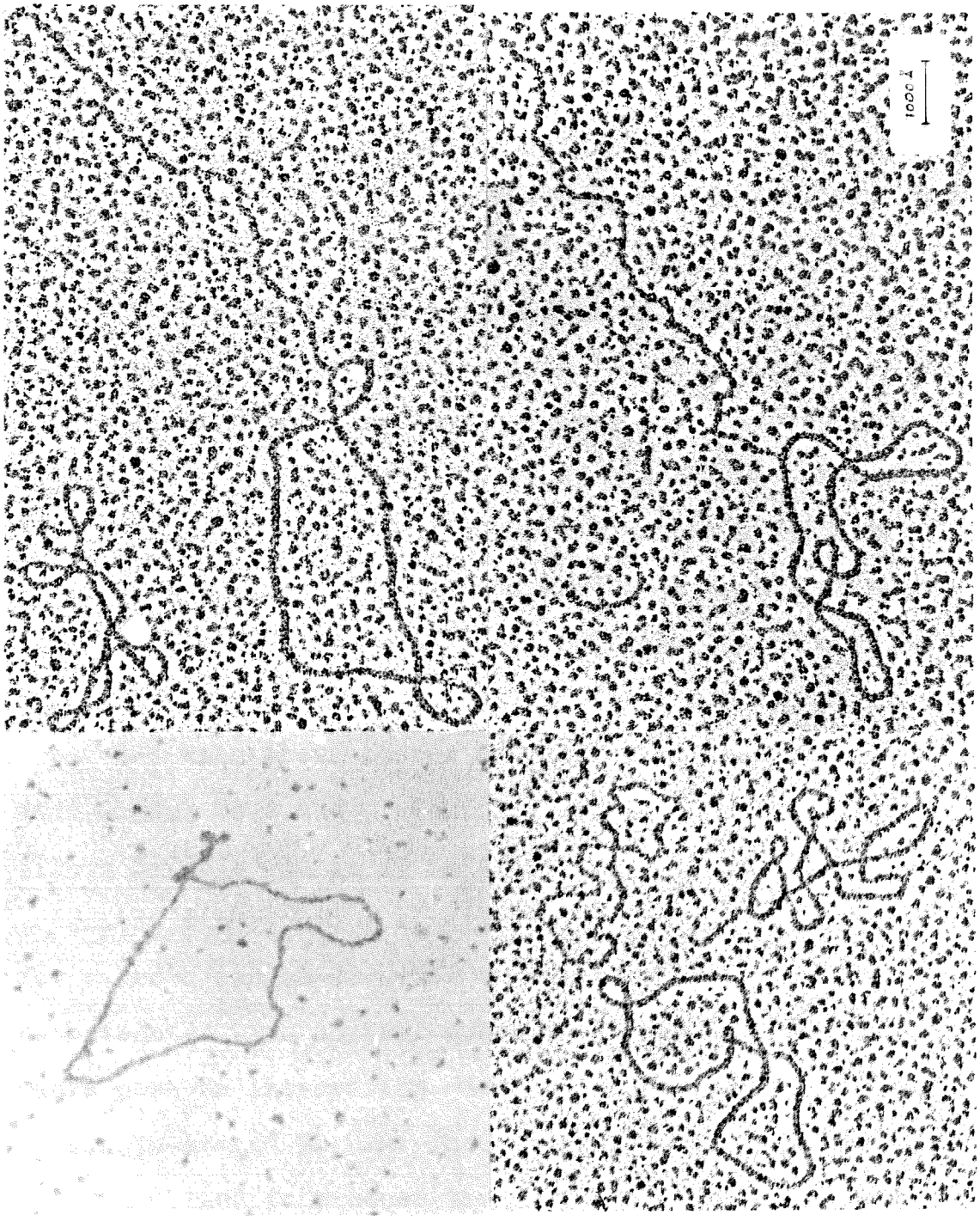
Plate 8-I

(a) ØX174 RFII with single-stranded DNA "bush". Mounted by the conventional technique and stained with uranyl acetate.

(b) Double-stranded ØX174 DNA mounted in formamide and shadowed with platinum-palladium.

(c) Single-strand circular ØX174 DNA mounted in formamide and shadowed with platinum-palladium.

(d, e, f) Double-stranded ØX174 DNA with single-stranded branch. The branches are shorter than whole single-stranded ØX174 circular DNA molecules. Micrographs (e) and (f) are 0.83 and 0.88 the length of whole single-stranded ØX174 DNA respectively.


1000 Å

(Plate 8-Ic). The longest branch observed was 0.97 the length of whole viral DNA. Figure 8-1 shows the distributions of single-stranded DNA branch lengths relative to whole single-stranded viral DNA. No significance is drawn from the fact that there is a peak in the distribution at 0.10 to 0.20. These molecules represent a fraction from a sucrose gradient, and there is the possibility of biasing the distribution.

Only one end of the single-stranded DNA branch was attached to the double-stranded RFII, and no replicating intermediates were seen that had both ends clearly attached to the RFII. This was true whether the DNA was mounted in the presence of formamide or only ammonium acetate. No single-stranded DNA branches were seen on RFI molecules (about 6% of the RFs were RFI). Multiple length circles of the RFIIs were also present and these are listed in Table 8-1 with their approximate number per cent. Catenated dimers were the predominant members in this class (Plate 8-IIa). It is of interest that a number of the catenated dimers had a single-stranded DNA branch. In a few cases a single-stranded circular molecule was seen to be catenated with an RFII molecule (Plate 8-IId). These facts pose an interesting question about the replication of single-strand ØX DNA: how is the linear single-stranded DNA closed to produce the circular viral DNA? Upon replication, nucleotides are added to the 3'-end of the viral strand with the displacement of the 5'-end from its

Fig. 8-1. Length distribution of the single-stranded DNA branch on replicating intermediates. The lengths were measured relative to mature viral single-stranded circular DNA. The replicating intermediates and viral DNA were mounted on the same grids, and were photographed simultaneously.

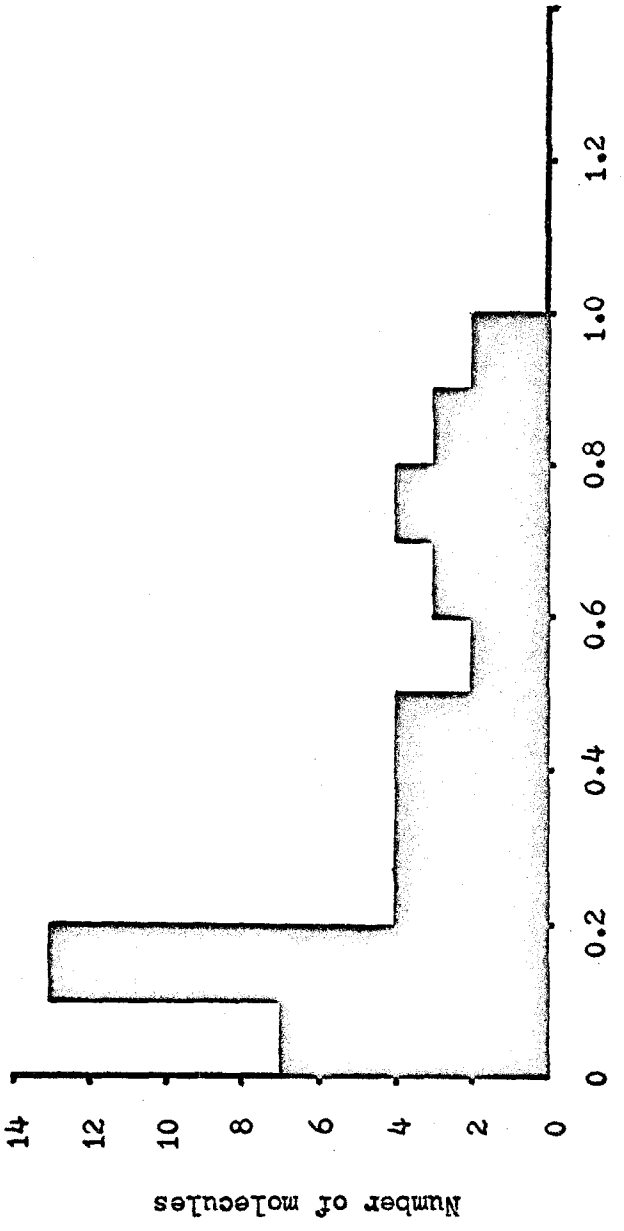

Number of molecules
14
12
10
8
6
4
2
0
0
0.2
0.4
0.6
0.8
1.0
1.2
Fractional length of single-stranded DNA branch on RFII
relative to single-stranded circular viral DNA

Table 8-1

| Different classes of ØX174 DNA from a sucrose gradient | Relative amounts in number per cent | | |
|---|---|---|---|
| | Preparation | | |
| | I | II | III |
| RFI | 53 | 6 | 6 |
| RFII | 25 | 61 | 70 |
| RFII with single-strand branch | 15 | 26 | 17 |
| RFII with double-strand branch | 0.5 | 0.5 | 0.5 |
| Catenated dimer* | 5 | 5 | 3 |
| Circular dimer | 0.5 | 0.5 | 2 |
| Linear catenated trimer | 0.5 | 0.5 | 0.5 |
| Higher multiple length circles** | 0.5 | 0.5 | 1 |

*In every preparation almost all (99%) of the multiple length circles were open (composed of RFII).

**Higher multiple length circles are catenated linear tetramers, circular dimers each catenated with a monomer, and circular trimers.

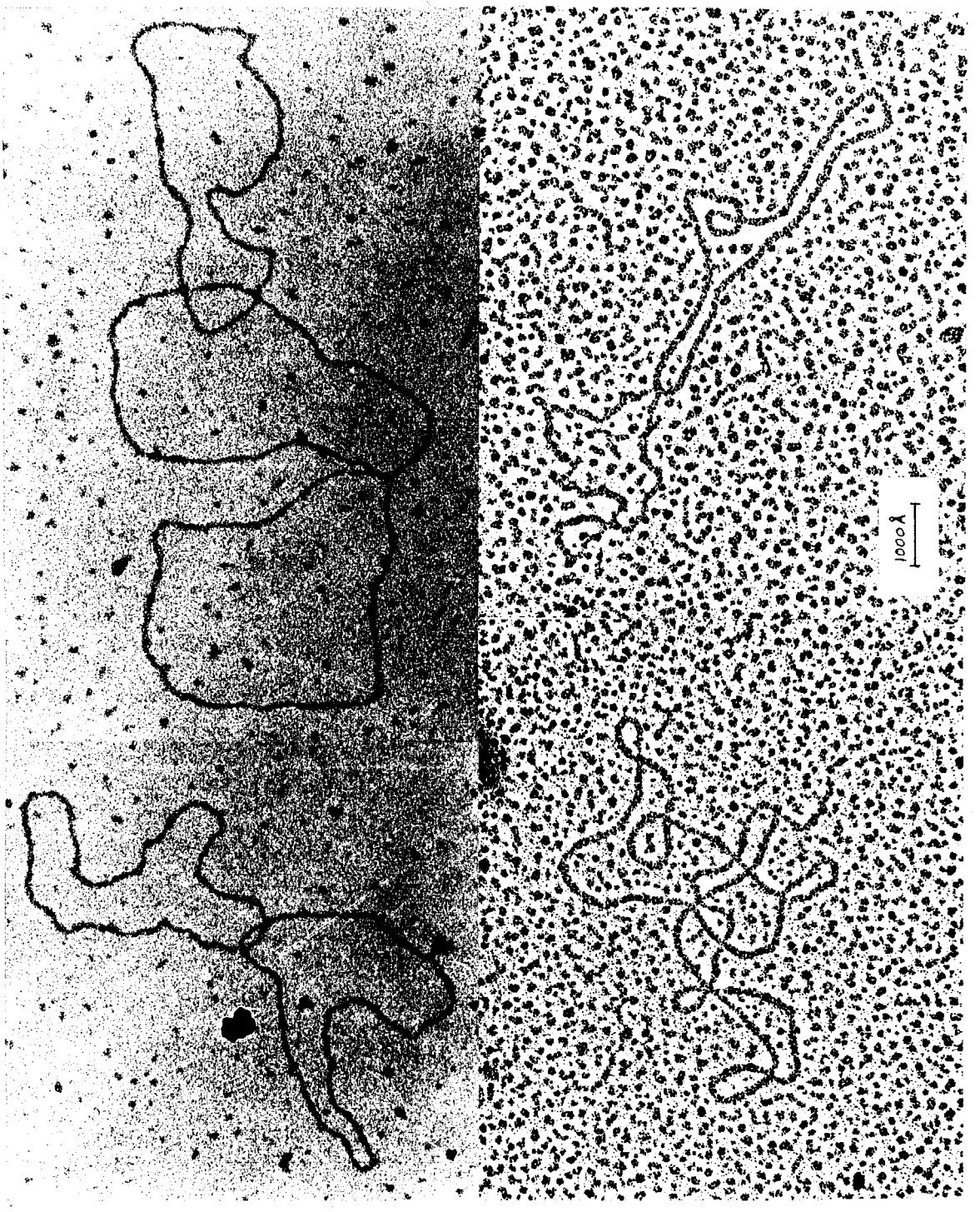
Plate 8-II

(a) Catenated dimer of RFII with a single-stranded DNA "bush". Mounted by the conventional technique and stained with uranyl acetate.

(b) Catenated trimer of RFII. Mounted as in (a).

(c) Catenated dimer of RFII with a single-stranded DNA branch. Mounted in 30% formamide and shadowed with platinum-palladium.

(d) Single-stranded circular molecule catenated with an RFII molecule.


1000 Å

complementary strand template. In order to cyclize, one might argue that the 5'-end is held on the RFII at the starting point for replication. Upon complete replication of the new viral strand, the two ends are in close proximity and in some manner are closed. This argument seems very unlikely since only one end of the growing viral strand is attached to the RFII template. One might then argue, however, that the 5'-end RFII linker was destroyed during the DNA isolation and handling. There is another line of reasoning, however, which completely rules out the linker theory. If the 5'-end was attached to its RFII template, then upon single-stranded DNA replication on a catenated dimer template, the newly replicated viral DNA would be catenated to its template after closure. This was never observed. In fact one might think that single-stranded closure should be made after its release from its template to avoid becoming catenated to its template.

## References

Davis, R. W. and N. Davidson (1968). PNAS, 60:243.

Denhardt, D. T., and R. L. Sinsheimer (1965). J. Mol. Biol., 12:647.

Hayashi, M., M. N. Hayashi, and S. Spiegelman (1963). Science, 140:1313.

Kleinschmidt, A. K. and R. K. Zahn (1959). Z. Naturforsch, 14b:770.

Komano, T., R. Knippers, and R. L. Sinsheimer (1968). PNAS, 59:911.

Knippers, R., T. Komano, and R. L. Sinsheimer (1968). PNAS, 59:577.

Knippers, Rolf, Aharon Razin, Ronald Davis, and Robert L. Sinsheimer, (1969). The Process of Infection with Bacteriophage ØX174 XXIX In Vivo Studies on the Synthesis of the Single-stranded DNA of ØX174 Bacteriophage, J. Mol. Biol., (in press).

Lindquist, B. H., and R. L. Sinsheimer (1967). J. Mol. Biol., 28:87.

Matsubara, K., A. Taketo, and Y. Takagi (1963). J. Biochem., Tokyo, 54:225.

Sinsheimer, R. L. (1961). J. Chim. Phys., 58:986.

Sinsheimer, R. L., B. Starman, C. Nagler, and S. Guthrie (1962). J. Mol. Biol., 4:142.

Sinsheimer, R. L., R. Knippers, and T. Komano, (1968).

Cold Spr. Harb. Symp. Quant. Biol.XXXIII, (in press).

Stone, A. B., (1967). Biochem. Biophys. Res. Commun. 26:247.

Yarus, M., and R. L. Sinsheimer, (1967). J. Virology, 1:135.

CHAPTER 9

# The Initiation Site for In Vitro RNA Synthesis on $T_7$ DNA

Ronald W. Davis and Norman Davidson

## 1. Introduction

Under appropriate conditions RNA can be synthesized *in vitro* from a DNA template by the enzyme RNA polymerase (Chamberlin & Berg, 1962; Weiss, 1960; Ochoa, Burma, Kroger, & Weill, 1961; Hurwitz, Bresler, & Diringer, 1960; Stevens, 1960; Weiss & Gladstone, 1959). An excellent review of *in vitro* RNA synthesis is given by Richardson (1969).

RNA polymerase can be reversibly dissociated into two components, the minimal enzyme and the $\sigma$ factor (Burgess, Travers, Dunn, & Bautz, 1969). The $\sigma$ factor seems to function at the level of initiation. It is released after the minimal enzyme-sigma complex binds to DNA and RNA synthesis is iniatiated. It is felt that for the minimal enzyme to reinitiate the synthesis of a new RNA chain it must first reform the minimal enzyme-sigma complex (Travers & Burgess, 1969). The minimal enzyme is composed of three principal types of polypeptide chain: $\alpha$, $\beta$, and $\beta'$. They are present in the enzyme in the ratios of 2 : 1 : 1 ' with respective molecular weights per each chain of 40,000, 155,000, and 165,000 daltons. There is one sigma factor polypeptide chain of molecular weight of 95,000 daltons associated with each minimal enzyme. Therefore, the total RNA polymerase enzyme is $\alpha_2\beta\beta'\sigma$ and has a molecular weight of 495,000

daltons (Travers & Burgess, 1969b; Burgess, 1969a). It may be of great significance that the sigma factor can be artificially released from the minimal enzyme during preparation. Also it does not bind to DNA and once released from the minimal enzyme it readily binds to glass walls and is irreversibly lost (Burgess, 1969b). Therefore, one of the primary sources of variability in the results of *in vitro* RNA synthesis may be due to differential loss of sigma.

It has been demonstrated (Bremer and Konrad, 1964; Geiduschek, Nakamoto, & Weiss, 1961) that there exists a stable intermediate for RNA synthesis consisting of an RNA-RNA polymerase-DNA complex. They also showed that the RNA was connected to the DNA by the RNA polymerase and denaturation of the polymerase (by SDS) released the RNA from its DNA template. However, there is an as yet unresolved dispute over this last point (Hayashi, 1965).

In the present study, electron-microscopic visualization of the RNA-RNA polymerase-DNA complex has been achieved. The complex is presumably located at the growing point of the RNA chain. The position of the complex at various times on $\lambda$ DNA and $T_7$ DNA was studied. In this manner, the region of the DNA template which is transcribed *in vitro* was mapped. The aim of the present study is to ascertain if there are unique sites along the DNA template at which RNA synthesis is initiated.

## 2. Materials and Methods

### (a) RNA polymerase

The RNA polymerase samples were prepared in Prof. James Bonner's laboratory by the procedure of Chamberlin & Berg (1962). The RNA polymerase used with λ DNA was a gift from Ron Jensen while the polymerase used with $T_7$ DNA was a gift from Grace Dahmus. The RNA polymerase from Grace Dahmus was assayed by Richard Hyman and was found to contain 2.5 X $10^3$ units of RNA polymerase / $A_{280}$ unit. One unit = 1mμmole of total NTP incorporated in 15 min at 35° C. under the salt conditions of So, Davie, Epstein, & Tissieres (1967) on 6γ of crab dAT template. From velocity sedimentation experiments on glycerol gradients, Richard Hyman estimated that the enzyme was about 40% pure.

### (b) RNA synthesis conditions

Although a large variety of synthesis conditions were used in this study, the conditions which seem to give the most selective initiation of RNA are as follows in the order of addition:

0.15 M KCl

0.01 M $MgCl_2$

0.05 M Tris-HCl pH 7.9 @ 37°C.

$10^{-3}$ M β-EtSH

10 γ/ml lambda or $T_7$ DNA

approx. 1γ/ml RNA polymerase (F4)

Ten min at 0°C. --- Binding Step

$4 \times 10^{-4}$ M ATP, GTP, & UTP

Ten min at 37.0°C. --- Initiation Step

$4 \times 10^{-4}$ M CTP

Incubated at 37.0°C. --- Propagation Step

(c) DNA

$\lambda_{b2b5c}$ and $\lambda_{c26}$ were grown as described in Davis and Davidson (1968). The DNA was extracted from the phage with phenol, followed by ether extraction of the phenol. $T_7$ DNA was a gift from Richard Hyman. It had been prepared in a similar fashion.

(d) Electron Microscopy

The DNA and the RNA-RNA polymerase-DNA complex were mounted for electron microscopy by the basic protein film technique and stained with uranyl acetate as described by Davis and Davidson (1968).

## 3. Results

### (a) Visualization of the RNA-RNA polymerase-DNA complex

RNA was synthesized on λ DNA using the RNA polymerase assay conditions of Chamberlin & Berg (1962). After 10 min a sample was removed and mounted by the aqueous basic protein film technique (Davis, Simon, & Davidson, 1969). Upon electron microscopic examination, many of the DNA molecules had what appeared to be a single-stranded polynucleotide bush attached to the duplex DNA (Plate 9-Ia & b). The bush appeared to be attached to the template at only one point and did not appear to distort the structure of the duplex DNA.

All evidence indicates that the bushes were RNA, synthesized by the DNA-dependant RNA polymerase. The bushes were not observed if any one of the necessary ingredients were omitted from the reaction mixture (DNA, RNA polymerase, $Mg^{++}$, ATP, GTP, CTP, and UTP) and they completely disappeared on treatment of the reaction mixture with RNase. Also, the bushes are not single-stranded DNA, resulting from the degradation of the DNA template, since the λ DNA with bushes had the same contour length as λ DNA without a visible bush. Furthermore, the bushes were released from the DNA by treatment with 0.01% SDS as expected from the results of Bremer and Konrad (1964),

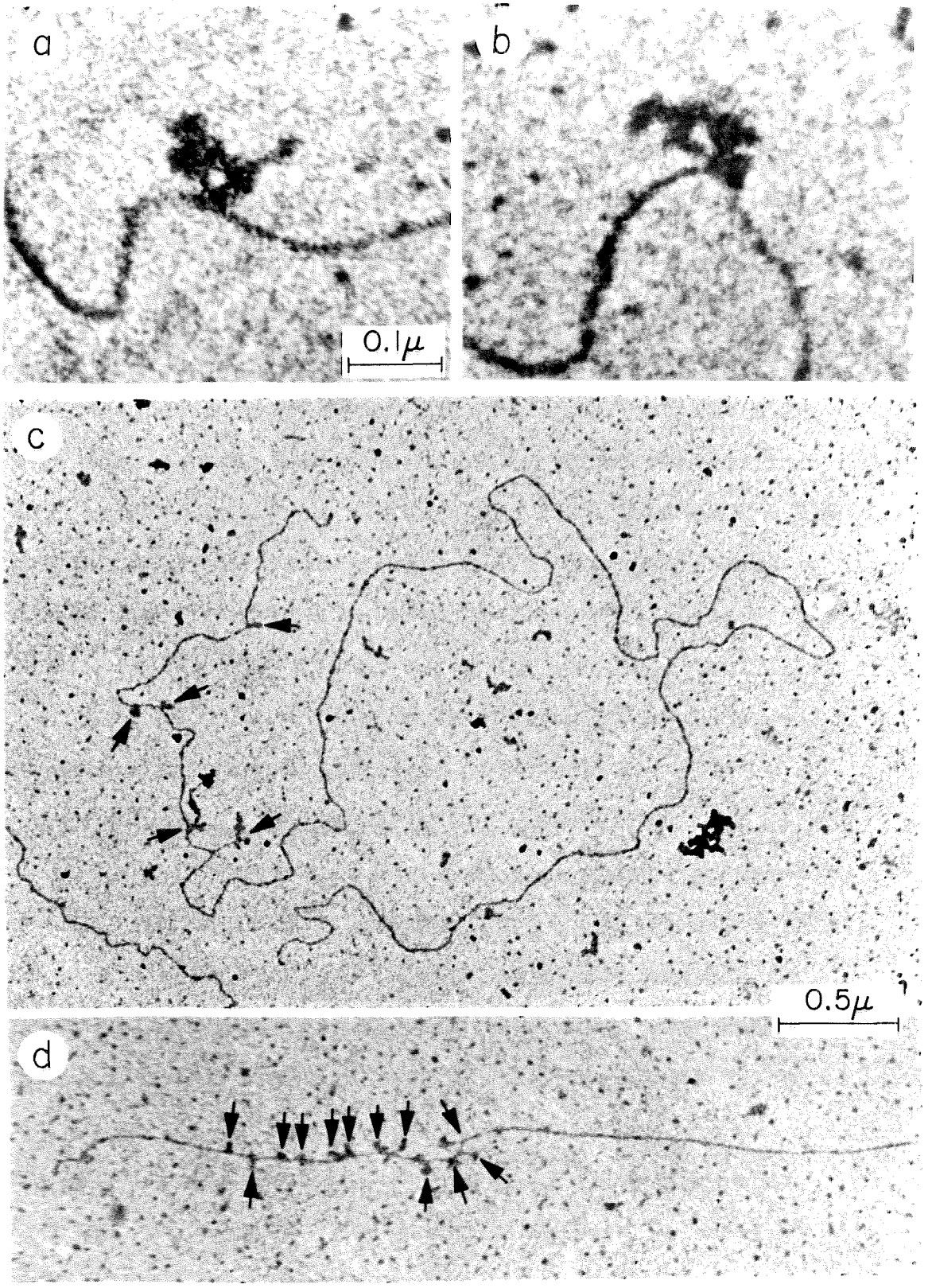
Plate 9-I.

(a) & (b) RNA bushes resulting from 5 min in vitro RNA synthesis on $T_7$ DNA. Note that the 2 bushes are about equal in size.

(c) In vitro RNA synthesis on $T_7$ DNA. Note that there are 5 RNA bushes on the DNA and that all of them are near one end. The synthesis reaction was carried out for 5 min.

(d) Same as (c) except the synthesis reaction was carried out for 2.5 min and at a slightly higher RNA polymerase/DNA ratio. Note that there are about 12 RNA bushes very near one end. The remaining part of the DNA not shown in the micrograph did not have any RNA bushes. The magnification of (c) and (d) are the same.

The RNA was synthesized under the conditions described in Materials and Methods, (b). The number of RNA polymerase molecules per DNA molecule was about 200. The DNA was mounted by the aqueous basic protein film technique and stained with uranyl acetate.


a
0.1μ
b
c
0.5μ
d

and Chamberlin (1969). Therefore, it is quite reasonable to assume that the bushes that are seen are indeed synthesized RNA.

All attempts by the basic protein film technique to see the RNA polymerase molecules attached to the DNA failed. The width of the DNA, with the collapsed protein film around it, is about 100 - 150 Å. Since the RNA polymerase is slightly smaller than this, one would not expect to be able to see it by the basic protein film technique unless the same amount of cytochrome c film that collapsed around the DNA also collapsed around the polymerase. Therefore, since RNA polymerase is one of the largest enzymes and since it can not be seen with the basic protein film technique, it is unlikely that this technique will be useful in the direct study of specific protein-DNA interactions.

(b) RNA synthesis on λ DNA

The most important question that can be answered by an electron microscopic analysis of *in vitro* RNA synthesis on a viral DNA template is whether or not the RNA is initiated at specific site(s) on the DNA. The bush positions of the attached RNA can be measured. If RNA synthesis is initiated at specific sites, then the RNA bush position should be unique. One difficulty with this approach is that the RNA moves along the DNA template during synthesis, and therefore, the RNA bush position

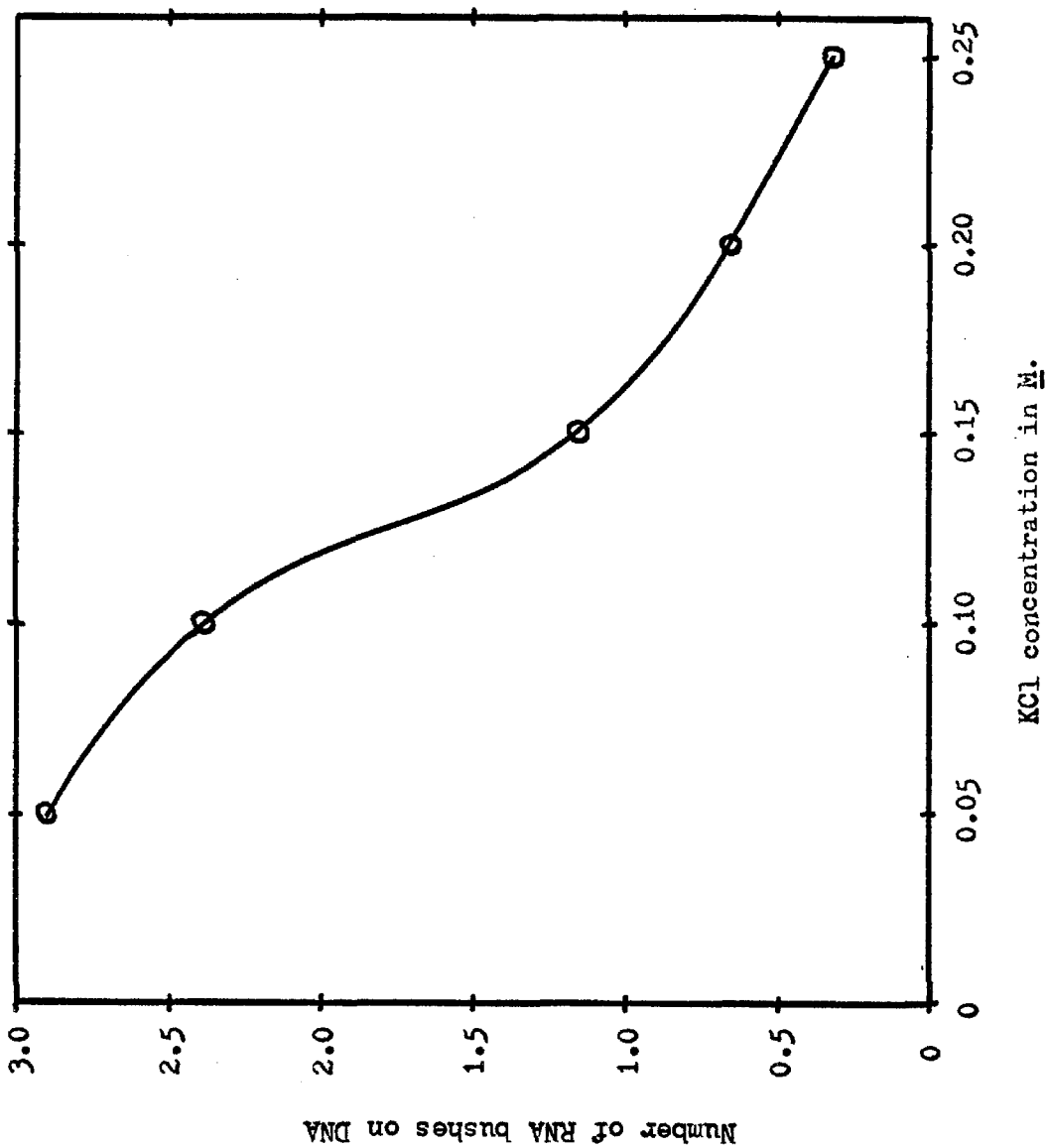
is a function of the amount of RNA in that bush.

The RNA bush positions on $\lambda_{b2b5c}$ DNA and $\lambda_{c26}$ DNA were measured after 5 and 10 min of RNA synthesis at 37° C. under the conditions of Chamberlin & Berg (1962). The amount of RNA polymerase used was adjusted so that there were about 5 observable RNAs per DNA. In every experiment, the RNA bush positions appeared to be more or less randomly located.

Since the Chamberlin & Berg conditions were originally selected to maximize the total amount of RNA synthesized, these conditions are probably also the least selective. Therefore, a synthesis condition was sought that would increase the initiation selectivity. The most obvious solution is to use a condition very close to in vivo conditions. This would mean omitting the $MnCl_2$, increasing the $MgCl_2$ to about 0.01 M, and adding KCl, (Lubin & Ennis, 1964). The most important change may be the increase in the ionic strength. Under higher salt conditions the presumed weak initiating sites may not be active. This idea was tested by synthesizing RNA for 5 and 10 min under various KCl concentrations (0.05, 0.10, 0.15, 0.20, and 0.25 M KCl). The results (and new synthesis conditions) are shown in Fig. 9-1. As can be seen fewer RNAs per DNA are synthesized at higher KCl concentrations. The distribution of the number of RNAs per DNA was analyzed at the five different KCl concentrations and

Fig. 9-1. Number of RNAs attached to λ DNA at various KCl concentrations during synthesis

The RNA synthesis mixture was made by adding to a small test tube 25μl of 0.5 M KCl, 71μl of $\lambda_{c26}$ DNA ($A_{260}$ = 0.883; 3.1γ), 10μl of 1M Tris, 51.5μl of $H_2O$, 2.5μl of 0.1 M β- mercaptoethanol, and 10μl of RNA polymerase stock solution (2γ of RNA polymerase). The reaction mixture was incubated for 20 min at 0° C. followed by the addition of 25μl of 0.1 M $MgCl_2$. 39μl of the reaction mixture was added to each of five tubes containing 5μl of $H_2O$, 0.5 M KCl, 1.0 M KCl, 1.5 M KCl, 2.0 M KCl respectively. Each tube was placed at 37° C. 1 min apart. 6μl of 3 X $10^{-3}$ M NTP (ATP, GTP, CTP, & UTP) was added to each tube. After 5 min of RNA synthesis at 37.0° C., 20μl of each reaction mixture was removed and placed in 1 ml of hyperphase (0.4 M $NH_4Ac$, 0.1 mg/ml cyt c, $10^{-3}$ M EDTA). The DNA was mounted for electron microscopy by the single drop method and stained with uranyl acetate.

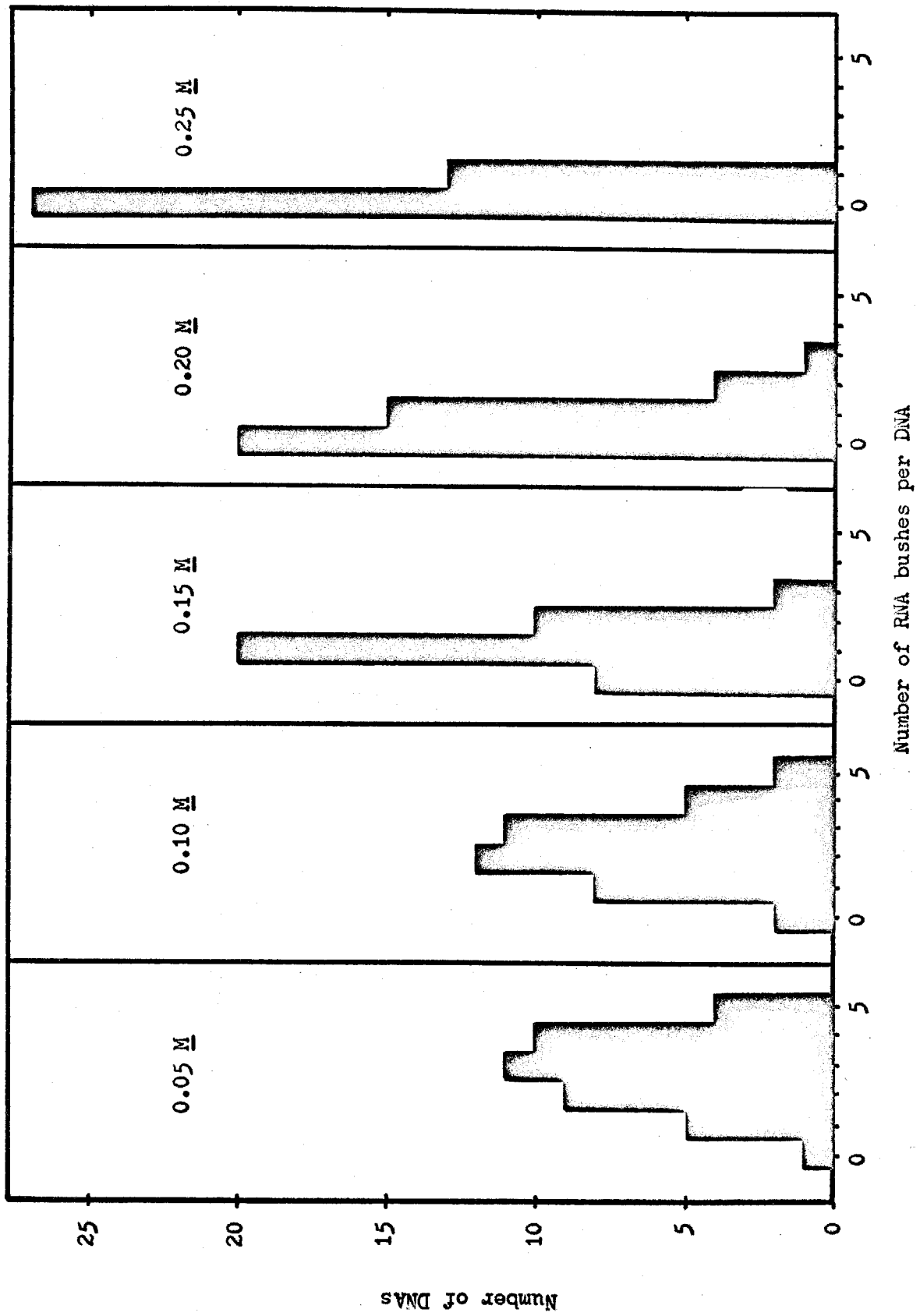

Number of RNA bushes on DNA
3.0
2.5
2.0
1.5
1.0
0.5
0
0
0.05
0.10
0.15
0.20
0.25
KCl concentration in M.

the results are plotted in Fig. 9-2. These distributions do not differ significantly from Poisson distributions. In this particular experiment, the stability of the RNA-RNA polymerase-DNA complex, under the electron microscopic mounting conditions (20 fold dilution into 0.05 M $NH_4Ac$, 0.1 mg/ml cytochrome c, $10^{-3}$ M EDTA), was tested. The complex was mounted on grids after storage at 0° C. for 30 secs, 2 hrs, and 2 days. There were no significant differences between the 30 sec and 2 hr storage. However, after storage for 2 days, the number of RNA-RNA polymerase-DNA complexes per DNA for RNA synthesized in 0.05 M KCl changed from 2.90 to 0.88 while for RNA synthesized in 0.15 M KCl the change was from 1.15 to 0.78. The reason for the difference in the stability after long storage is not understood except that it might result from slightly different levels of RNase in the samples. However, it can be concluded that the slight variability in the length of storage ($\pm$ 1 min) does not affect the results reported here. A single RNase cleavage of the RNA in the RNA-RNA polymerase-DNA complex is sufficient to release the bulk of the RNA from the DNA template. The remaining fragment of RNA may be invisible under the mounting conditions used here.

A study was again made into the selectivity of initiation of RNA under these new synthesis conditions by analyzing RNA bush positions. The KCl concentration was

Fig. 9-2. The distribution of the number of RNA bushes on the λ DNA template at various KCl concentrations during synthesis.

The RNA synthesis mixtures at various salt concentrations described in Fig. 9-1 were also assayed for the number of RNA bushes on each λ DNA template. The KCl concentration present during the RNA synthesis is shown at the top of each histogram. A total of 40 DNA molecules were picked at random and scored as to the number of RNAs on each.


0.05 M
0.10 M
0.15 M
0.20 M
0.25 M
Number of DNAs
0
5
10
15
20
25
0
5
0
5
0
5
0
5
0
5
Number of RNA bushes per DNA

0.15 M for most investigations. It was discovered that no unique RNA bush positions were present after 5 or 10 min of synthesis. However, there were certain regions on the $\lambda_{c26}$ DNA template where no RNA bushes were seen either after 5 min or after 10 min of synthesis. These regions are roughly from 0 to 0.1, 0.4 to 0.6, and 0.9 to 1.0 fractional lengths of $\lambda^+$ DNA. Therefore, although no unique positions for *in vitro* RNA synthesis on λ DNA could be found, certain regions on the DNA were excluded from acting as a template for 5 and 10 min RNA synthesis.

(c) RNA synthesis on $T_7$ DNA

Although there are many explanations as to why *in vitro* RNA synthesis on λ DNA did not give uniquely positioned RNA bushes, a major cause may be the use of λ DNA as the template. It is known that λ DNA is a poor template for RNA synthesis (Jones & Berg, 1966). Therefore, a better template for this study is $T_7$ DNA. This DNA is a much better template for RNA synthesis than λ DNA (Jones & Berg, 1966; Stead & Jones, 1967). Also, it is known that only one of the $T_7$ DNA strands is used as a template for *in vivo* and *in vitro* RNA synthesis (Szybalski, 1969; Summers & Szybalski, 1968).

It was also learned from Chamberlin (1969) that the initiation step in the synthetic reaction is very slow. It has been assumed thus far that initiation was very fast, and that propagation was relatively slow (2-6 bases

per sec) (Bremer & Konrad, 1964; Richardson, 1967). The slow initiation step could explain the absence of unique RNA bush positions. The bushes will not be uniquely positioned if the RNAs start propagating at different times, even though they started at unique sites along the DNA. Therefore, the following preinitiation step was adopted. The DNA and RNA polymerase were first mixed in the appropriate salt solution. Ten min were allowed for the binding reaction. Then ATP, GTP, and UTP were added and 15 min were allowed for the initiation step. The propagation step was started by the addition of CTP.

Using this modification of the synthesis reaction, RNA was synthesized on $T_7$ DNA. Synthesis samples were removed after 5 and 10 min. Approximately 5 RNA bushes were seen per DNA. The striking result of this experiment was that after 5 min of RNA synthesis all five of the RNA bushes were generally seen on only one half of the DNA (Plate 9-Ic). The results were not so clear after 10 min. The synthesis was repeated at a slightly higher RNA polymerase concentration and with an additional sample being removed after 2.5 min of RNA synthesis. The results for the 5 and 10 min samples were about the same as in the previous synthesis except more RNA bushes were seen. However, the sample removed after 2.5 min showed extreme clustering of the RNA bushes near one end. Plate 9-Id shows 12 RNA bushes on one end. The positions of the

bushes ranged from 0.062 to 0.140 fractional lengths of whole $T_7$ DNA. Therefore, it appears that under these conditions, all of the initiation sites for in vitro RNA synthesis on $T_7$ DNA are very near one end.

In these synthesis reactions there were observed only about 5 RNA bushes per DNA, however, approximately 200 RNA polymerase molecules were used per $T_7$ DNA. This calculation is made from the $A_{280}$ of the RNA polymerase stock solution (using $A_{280}^{1\%} = 6.6$, Burgess, et al, 1969) and from the fact that 40% of the $A_{280}$ absorbing material in the polymerase stock solution sediments in a glycerol gradient with the correct sedimentation coefficient and with the peak of the RNA polymerase activity (Hyman, 1969). Although this calculation of 200 RNA polymerase molecules per DNA is at best only an approximation, it is significantly different from the observation that only 5 RNAs are made per DNA (1 RNA per 40 RNA polymerase molecules).

In an effort to clarify this discrepancy and to investigate further the selective initiation of in vitro RNA on $T_7$ DNA, synthesis reactions were carried out at a polymerase molecules/DNA molecule ratio of 40. Samples were removed at 1, 2.5, and 5 min after the start of propagation. The samples were mounted for electron microscopy (Plate 9-II). The number of RNAs per DNA and the RNA bush positions were measured. It was found that there was approximately 1 RNA bush per DNA which again gives 1 RNA

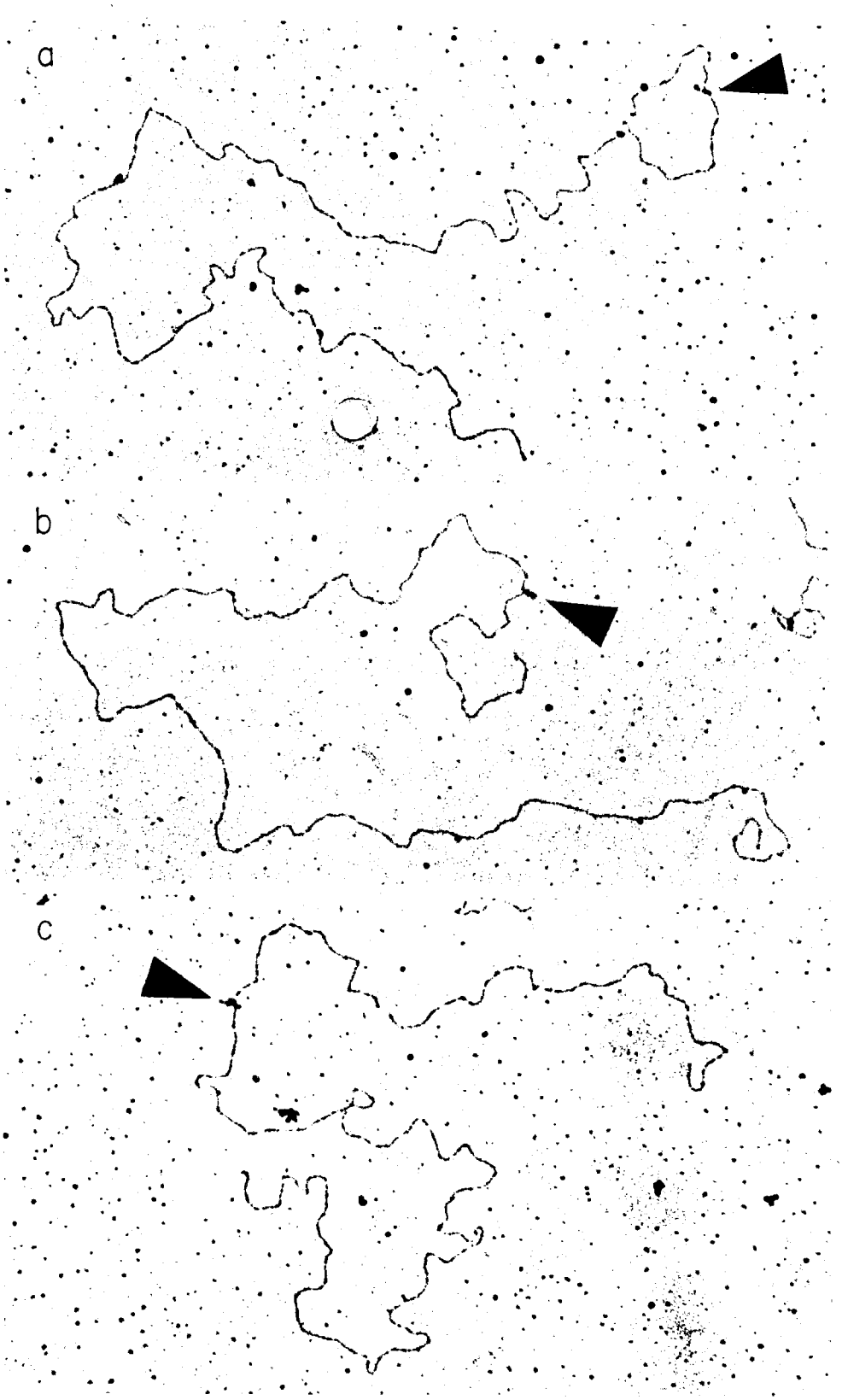
Plate 9-II. RNA synthesis on $T_7$ DNA

(a) RNA synthesized for 1.0 min.

(b) RNA synthesized for 2.5 min.

(c) RNA synthesized for 5.0 min.

The arrow denotes the position of the RNA bush on each micrograph. The RNA was synthesized as described in the text. The number of RNA polymerase molecules per DNA molecule in the reaction mixture was about 40. The synthesis samples were mounted for electron microscopy by the aqueous basic protein film technique and stained with uranyl acetate.

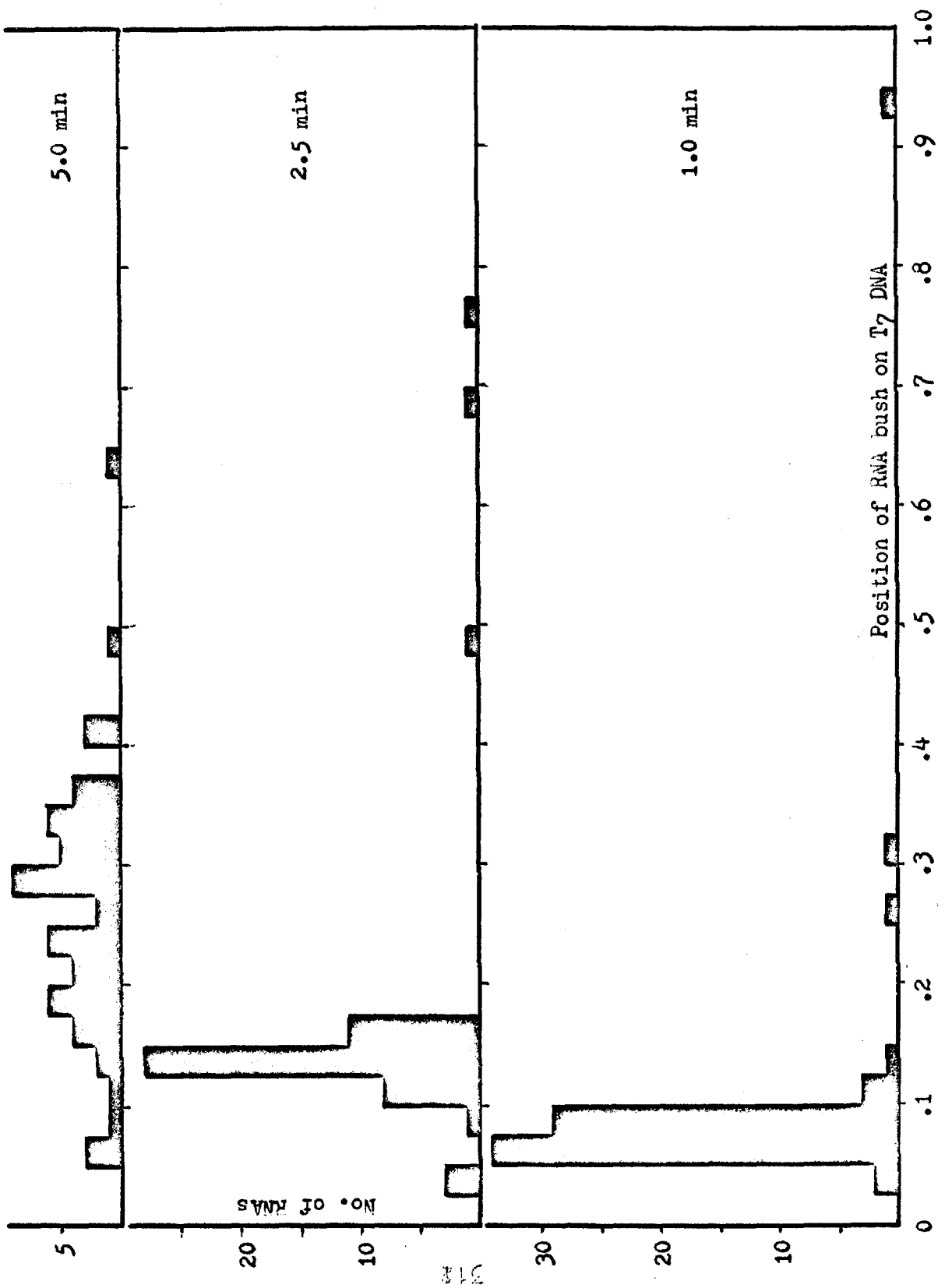

a
b
c

per 40 RNA polymerase molecules. Since there were a few DNAs which did not have an RNA bush, the concentration of RNA polymerase is probably limiting, and if the polymerase concentration were further reduced proportionately fewer RNA bushes would be seen. As shown in Fig. 9-3, unique RNA bush positions were found after 1 min and 2.5 min, while after 5 min there was a considerable spread of the bush positions. All of the RNAs appear to start near one end and propagate toward the other end. The rate of propagation was calculated from the distance the RNA bush positions moved along the DNA and was found to be 45 bases per sec between 1 and 2.5 min of synthesis and 40 bases per sec between 2.5 and 5 min of synthesis. From the appearance of the histograms in Fig. 9-3, it would seem that the RNA molecules are quite synchronized at the start of propagation. The synchronization appears to be progressively lost and by 5 min of synthesis there is little evidence of synchronization. From the width of the distribution of RNA bush positions after 1 min of synthesis, and the rate at which this distribution spreads, it would appear that all of the RNAs synthesized on $T_7$ DNA *in vitro* are initiated at the same point near the end of the DNA. This would imply that there is only one RNA initiating site on $T_7$ DNA.

Because RNA initiation has in the past been only vaguely understood, little effort has been made to study

Fig. 9-3. Physical map of RNA attachment site on $T_7$ DNA at various propagation times.

The positions of the RNA bushes on $T_7$ DNA at three propagation times are shown. There was approximately 1 RNA synthesized per DNA. Synthesis was carried out in 0.15 M KCl, 0.01 M $MgCl_2$, 0.05 M Tris, pH 7.9, $10^{-3}$ M $\beta$- EtSH, 10 $\gamma$/ml BSA, 10 $\gamma$/ml $T_7$ DNA, 5$\gamma$/ml RNA polymerase, and 4 X $10^{-4}$ M NTP. Preinitiation, by omitting the CTP, was used to synchronize the propagation of RNA.


5.0 min
2.5 min
1.0 min
No. of RNAs
Position of RNA bush on T7 DNA
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
5
20
10
30
20
10

the termination of RNA synthesis. The experimental approach given here would seem ideal for the study of RNA termination. However, few conclusions can be made about termination. It has been consistently observed that about 50% of the RNA is released from the DNA after 5 min of synthesis. Released RNA can not be positively identified for shorter synthesis times because of its small size. After 5 min of synthesis some RNA is located at 0.3 fractional $T_7$ DNA units from the end. Therefore, one conclusion might be that there are no termination sites between 0 and 0.3 units from the end and that the released RNA is solely due to RNase activity. However, RNA could be uniquely released without release of the polymerase. The polymerase could then continue synthesis along the DNA.

## 4. Discussion

In this study it was discovered that under appropriate conditions selective initiation sites for *in vitro* RNA synthesis on $T_7$ DNA exist. The RNA synthesis was carried out in 0.15 M KCl, 0.01 M $MgCl_2$, 0.05 M Tris, and $4 \times 10^{-4}$ M NTP. A binding step at $0^{\circ}$ C. with just DNA and polymerase in the above salt was followed by a preinitiation step in which ATP, GTP, and UTP were added. The propagation was started by the addition of CTP. Under these conditions the propagation of RNA was synchronized and thus the RNA bush positions were uniquely located. From the analysis of the bush positions, it appears that there is only one region near one end in which RNA can be initiated on $T_7$ DNA. Since this region of DNA seems to be less than 1% of the $T_7$ DNA it would appear that there is only one initiation site or several tandemly arranged initiation sites. The fact that several RNAs can be synthesized on the $T_7$ DNA at the same time does not indicate several initiation sites since this could be the result of sequential binding, initiation, and propagation from a single site on the DNA.

The conclusion that $T_7$ DNA has only one *in vitro* RNA initiation site is not in agreement with all previously published determinations. It has generally been concluded from RNA polymerase binding studies and from the number of

RNA molecules initiated, that $T_7$ DNA has about 40 *in vitro* RNA initiating sites. However, in these calculations it was assumed that all of the polymerase molecules in their enzyme preparation are active, that polymerase binds only to initiating sites, and that there is only one RNA initiated per initiation site.

The rate of propagation of the RNA was measured to be about 45 bases/sec. This value is considerably higher than the previously measured values. At saturating concentration of nucleoside triphosphates and in low salt the RNA growth rate is 4 bases/sec on $T_4$ DNA (Bremer, 1967); while in 0.2 M KCl the rate has been measured to be 16 bases/sec. However, in these determinations, synchronized initiation was not employed, and therefore, these growth rates are underestimates. The synthesis rate observed here is in good agreement with the rate observed *in vivo* (43 bases/ sec at 37.5° C.) (Manor, Goodman, & Stent, 1969). The conclusions given here, that there is one initiation site, that several RNAs can be sequentially initiated from this site, and that the rate of propagation is 45 bases/sec is not inconsistent with any previously published data although it is inconsistent with the conclusions drawn from these data.

In these experiments the start of the propagation of RNA was synchronized by a preinitiation step. However, synchronization of propagation was lost after 5 min. Since

the bound RNA/DNA ratio did not significantly change with time, this loss of synchronization can be explained by a differential rate of RNA propagation or by the release of RNA bushes and reinitiation of new RNAs at the RNA initiation site.

It was observed in these experiments that there were about as many free RNA bushes as there were DNA bound RNA bushes after 5 min of synthesis. These free RNA bushes could have occured by RNase cleavage of the bound RNA, by denaturation of the polymerase and random release of the bound RNA, or by selective release at unique sites along the DNA.

There were many experimental difficulties encountered in this set of experiments. These difficulties should be borne in mind by anyone in attempting a more refined interpretation of the present experiments or in planning to do new experiments. The conditions and ingredients used for RNA synthesis are not favorable for good electron microscope experiments: the molecules tend to be tangled, the contrast is poor, and the RNA bushes are highly clumped. There was some degradation of both DNA and RNA. This is presumably the result of nucleases in the RNA polymerase stock solutions. However, this degradation was significant only for long synthesis reaction times (20 min or more) and did not effect the results for the short incubations (1 - 5 min.).

In light of what has been learned about RNA synthesis on $T_7$ DNA, it is not surprising that RNA synthesis on λ DNA did not show uniquely positioned RNA bushes. RNA synthesis on λ DNA was carried out for 5 to 10 min and the preinitiation step was not used. Shorter synthesis times were not studied since RNA bushes could not be visualized. The lack of bushes at short synthesis times (less than 5 min) is presumably due to the slow initiation of RNA. Therefore, it is quite possible that uniquely positioned RNA bushes would be found on λ DNA if preinitiation and short synthesis times are used.

## References

Bremer, H., (1967). Mol. Gen. Genet., 99, 362.

Bremer, H. & M. W. Konrad, (1964). Proc. Nat. Acad. Sci., Wash., 51, 801.

Burgess, R. R. (1969a). Thesis, Harvard University.

Burgess, R. R. (1969b). Personal communication.

Burgess, R. R., A. A. Travers, J. J. Dunn, & E. K. F. Bautz, (1969). Nature, 221, 43

Chamberlin, M., (1969). Personal communication.

Chamberlin, M. & P. Berg, (1962). Proc. Nat. Acad. Sci., Wash., 48,81.

Geiduschek, E. P., T. Nakamoto, and S. B. Weiss, (1961). Proc. Nat. Acad. Sci., Wash., 47,1405.

Hayashi, M., (1965). Proc. Nat. Acad. Sci., Wash., 54, 1736

Hurwitz, J., A. Bresler, & R. Diringer, (1960). Biochem. Biophys. Res. Commun., 3,15.

Hyman, Richard, (1969). Personal communication.

Jones, O. W. & P. Berg (1966). J. Mol. Biol. 22, 199.

Lubin, M. & H. L. Ennis, (1964). Biochem. Biophys. Acta, 80, 614.

Manor, H., D. Goodman, & G. S. Stent (1969). J. Mol. Biol., 39,1.

Ochoa, S., D. P. Burma, H. Kroger, & J. D. Weill, (1961). Proc. Nat. Acad. Sci., Wash., 47,670.

Richardson, J. P., (1969). Progress in Nucleic Acid Research and Molecular Biology., 9,75.

So, A. G., E. W. Davie, R. Epstein, & Tissieres, (1967). Proc. Nat. Acad. Sci., Wash., 58,1739.

Stead, N. W. & O. W. Jones, (1967). J. Mol Biol., 26,131.

Stevens, A., (1960). Biochem. Biophys. Res. Commun., 3,92.

Summers, W. C. & W. Szybalski, (1968). Virology, 34,9.

Szybalski, W., (1969). Personal communication.

Travers, A. A. & R. R. Burgess,(1969a). To be published.

Travers, A. A. & R. R. Burgess, (1969b). Nature, 222,537.

Weiss, S. B. (1960). Proc. Nat. Acad. Sci., Wash., 46,1020.

Weiss, S. B. & Gladstone, (1959). J. Am. Chem. Soc., 81, 4118.