

I

THE RELATIVE HOMOGENEITY OF MICROBIAL  
DESOXYRIBONUCLEIC ACID

II

THE MOLECULAR ARRANGEMENT OF THE CONSERVED  
SUBUNITS OF THE DESOXYRIBONUCLEIC ACID  
OF ESCHERICHIA COLI

Thesis by

Ronald Rolfe

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology  
Pasadena, California

1961

## ABSTRACT

### I

The buoyant density of DNA in a CsCl-density gradient has been shown to depend on its nucleotide composition. The linear relationship has been used to study the distribution of nucleotide compositions among the DNA molecules isolated from a single microbial species. Each microbial DNA has been shown to be unusually homogeneous relative to the range of compositions found for DNA isolated from different species. The relevance of these findings to current views of the function of DNA in biological systems is discussed.

### II

The molecular arrangement of the conserved subunits of E. coli DNA has been investigated by examining molecular fragments of hybrid  $^{13}\text{C}$ ,  $^{15}\text{N}$  E. coli DNA. Two extreme models for the arrangement of the subunits, the side-to-side model, and the end-to-end model, were considered. Predictions regarding the CsCl-density-gradient distribution for fragments of hybrid DNA were developed for each model. These predictions were based on a theoretical analysis, but utilized pertinent experimental data obtained from a study of unlabelled E. coli DNA. Comparison of theory with experiment indicated that the end-to-end model is incorrect, and set an upper limit to the amount of fully labelled fragments released when hybrid DNA is sonicated.

## ACKNOWLEDGEMENTS

The guidance and friendship of my thesis advisor, Professor Matthew Meselson, and the helpful advice of my fellowship sponsor, Professor Linus Pauling, have been deeply appreciated.

Professors Max Delbruck, Robert Sinsheimer and Jean Weigle gave aid and helpful criticism of my research.

The National Institutes of Health provided financial support during the tenure of this work.

## TABLE OF CONTENTS

	<u>Page</u>
I THE RELATIVE HOMOGENEITY OF MICROBIAL DNA . . . . .	1
II THE MOLECULAR ARRANGEMENT OF THE CONSERVED SUBUNITS OF <u>E. COLI</u> DNA . . . . .	6
A Introduction . . . . .	6
1 Statement of the Problem . . . . .	6
2 The Initial Approach to the Problem . . . . .	8
3 A Second Approach to the Problem . . . . .	9
4 Outline of Experiments . . . . .	12
B Experimental Part . . . . .	16
1 Isolation of DNA . . . . .	16
2 Transfer Experiments . . . . .	16
3 Fragmentation of DNA . . . . .	17
4 Electron Microscopy . . . . .	17
5 Ultracentrifuge Studies . . . . .	18
Photography . . . . .	18
Sedimentation-Velocity Studies . . . . .	19
Density-Gradient Studies . . . . .	20
C Results . . . . .	22
1 Properties of sonic fragments of unlabelled DNA . . . . .	22
The calculation of $\sigma_1^{-2}$ for the end-to-end model . . . . .	24
The calculation of $\sigma_{M_n}^{-2}$ from sedimentation-velocity data . . . . .	25
The calculation of $\sigma_{GC}^{-2}$ from density-gradient data . . . . .	26
The calculation of $\sigma_T^{-2}$ expected for the two models . . . . .	26
The determination of $\sigma_T^{-2}$ from half bands . . . . .	26
2 Properties of sonic fragments of transfer DNA . . . . .	27
Predictions for the end-to-end model . . . . .	29
Predictions for the side-to-side model . . . . .	29
Characteristics of the experimental distributions . . . . .	30
D Summary . . . . .	32

APPENDIX 1	
Calculation of equilibrium distributions for fragments of hybrid DNA from models for the arrangement of the subunits . . . . .	33
APPENDIX 2	
The relationship between the variance of the equilibrium distribution of fragments of hybrid DNA, and the variance of the distributions of nucleotides and heavy isotopes among the molecules . . . . .	43
APPENDIX 3	
Calculation of the time required to reach equilibrium for CsCl density-gradient centrifugation of DNA . . . . .	45
APPENDIX 4	
The determination of molecular weight for samples of DNA . . . . .	47
APPENDIX 5	
Symmetry properties of DNA bands . . . . .	51
TABLES . . . . .	54
FIGURES . . . . .	58
REFERENCES . . . . .	89
PROPOSITIONS . . . . .	91

Reprinted from the Proceedings of the NATIONAL ACADEMY OF SCIENCES  
Vol. 45, No. 7, pp. 1039-1043, July, 1959.

# I

## *THE RELATIVE HOMOGENEITY OF MICROBIAL DNA\**

BY RONALD ROLFE† AND MATTHEW MESELSON

GATES AND CRELLIN LABORATORIES OF CHEMISTRY‡ AND NORMAN W. CHURCH LABORATORY OF  
CHEMICAL BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA

*Communicated by Linus Pauling, May 6, 1959*

Density-gradient centrifugation has revealed that the population of DNA molecules from *E. coli* is relatively homogeneous with respect to buoyant density in a solution of cesium chloride.<sup>1</sup> Because of the notably small apparent atomic volume of nitrogen in aqueous solutions of organic compounds<sup>2</sup> and because the guanine-cytosine base pair is more rich in nitrogen than is the adenine-thymine pair, it was considered that the density homogeneity among *E. coli* DNA molecules might reflect a high degree of homogeneity with respect to base composition. To investigate this possibility, an examination has been made of the relationship between buoyant density in cesium chloride solution and base composition of DNA from various sources.



Equation (1) was obtained from determinations of the mean composition and mean buoyant density of DNA from several species of bacteria but we may now apply it to the population of DNA molecules from a single species. In particular,

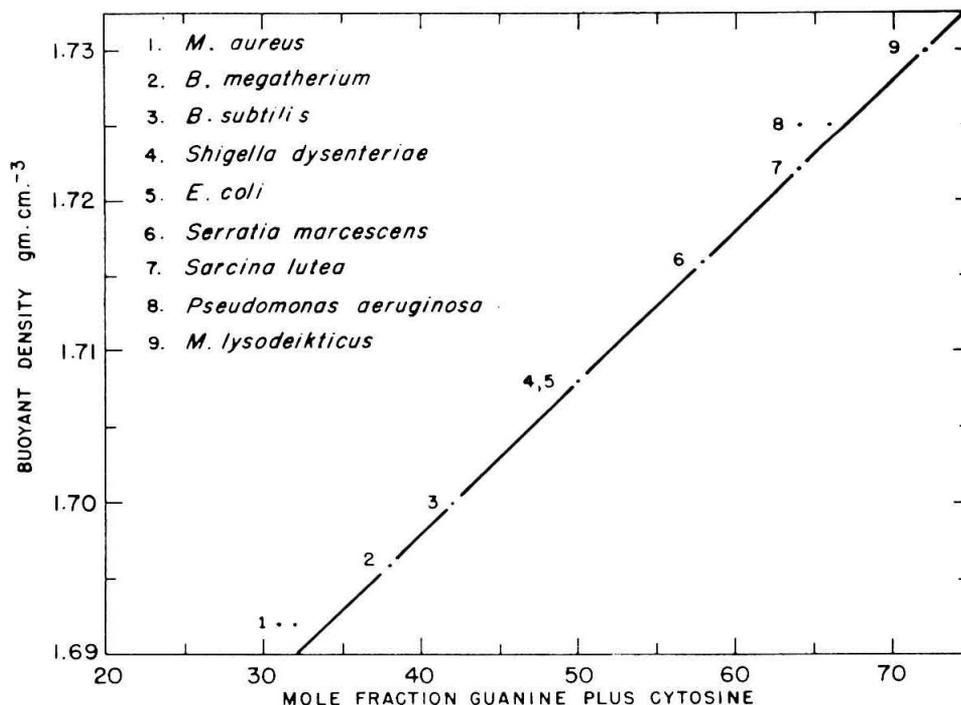


FIG. 2.—The density-composition relationship for DNA. Mean buoyant density at 25°C. ( $P_{25}^{\circ}$ ) is plotted against mean molar fraction of guanine plus cytosine (GC) for the DNA from nine bacterial species. Where GC is variable among the studied strains of a given species, the range of variation is indicated by a pair of points. The base composition given for *Shigella dysenteriae* is that of *Shigella paradysenteriae*.

we shall compute for each species an upper bound on the standard deviation  $\sigma_{GC}$  of the distribution of the guanine-cytosine base pair over the population of DNA molecules. From equation (1) the upper bound on  $\sigma_{GC}$  is given as

$$\sigma_{GC \text{ max}} = 10.0\sigma_p$$

where  $\sigma_p$  is the standard deviation of the DNA distribution in the cesium chloride density gradient. This computation requires for its validity that no DNA band be seriously narrowed by intermolecular aggregation and that each band comprise a representative sample of the total DNA of the corresponding bacterial species. Aggregation is unlikely in view of the absence of aggregation in a cesium chloride density gradient between molecules of *E. coli* DNA differing only in their content of the heavy isotope  $N^{15}$ . It is very likely that the banded DNA is representative of the total DNA because of the close agreement between the mean GC content determined chromatographically and that calculated from buoyant density by means of equation (1). It should be pointed out that the actual value of  $\sigma_{GC}$  may lie considerably below the calculated upper bound because thermal motion of the DNA molecules contributes significantly to the bandwidth.

The DNA from each bacterial species investigated forms a band in the density gradient with  $\sigma_p$  in no case greater than  $0.003 \text{ gm. cm}^{-3}$ . The corresponding upper bound on the standard deviation  $\sigma_{GC}$  of the molecular content of guanine plus cytosine is therefore in no case greater than 0.03.<sup>5</sup> It is remarkable that the standard deviation of guanine-cytosine content within the molecular population of any one bacterial species covers less than one tenth of the range over which the mean guanine-cytosine content varies among the various species.

If we assume for purposes of discussion that much hereditary information is common to the various bacterial species and that DNA is a carrier of genetic information, then our finding that few, if any, DNA molecules possess compositions common to the various species argues against the conception that the complete nucleotide sequence could in principle be deduced from other hereditary characteristics according to a universal code. Instead, it may be that only certain features of the nucleotide sequence are genetically significant so that extensive modification of nucleotide composition need not result in any other genetic alteration. Or it may be that the detailed relation between nucleotide sequence and genetic specificity is itself a species characteristic.

As this investigation was being completed, we learned that Sueoka, Marmur, and Doty have independently arrived at a relation between DNA composition and buoyant density which is in good agreement with the relation reported here. We wish to thank these authors for communicating their findings to us before publication.

\* Aided by a grant from the National Institutes of Health.

† Postdoctoral Fellow of the U. S. Public Health Service.

‡ Contribution No. 2460.

<sup>1</sup> Meselson, M., and F. W. Stahl, these PROCEEDINGS, **44**, 671 (1958).

<sup>2</sup> Traube, J., *Samml. Chem. u. Chem.-Tech. Vorträge*, **4**, 255 (1899).

<sup>3</sup> Lee, K. Y., R. Wahl, and E. Barbu, *Ann. Inst. Pasteur*, **91**, 212 (1956). Base compositions are those given by Lee *et al.* except for *E. coli* which was not reported by them. Data for *E. coli* is from Smith, J. D., and G. R. Wyatt, *Biochem. J.*, **49**, 144 (1951).

<sup>4</sup> The mean buoyant densities of calf thymus, salmon sperm, and human leukocyte DNA all lie close to the values calculated from equation (1). However, the density distribution of the DNA from each of these vertebrate sources exhibits marked skewness toward higher density (M. Meselson, Thesis, California Institute of Technology, 1957). This skewness might reflect an uneven distribution of GC among the molecules of DNA or may be due to an uneven distribution of the rare base 5-methyl cytosine which occurs in small amounts in the DNA from these three sources.

The density composition relationship (1) is not valid for every DNA which has been examined. Density values considerably higher than those calculated from (1) are observed for heat denatured DNA from *E. coli*, calf thymus, and salmon sperm,<sup>1</sup> for T2 and T4 phage DNA in which cytosine is replaced by 5-hydroxymethyl cytosine or glucosylated 5-hydroxymethyl cytosine, and for DNA from the phage  $\phi$ X 174, found to have an unusual structure and composition [Sinsheimer, R. L., *J. Molec. Biol.* (in press)]. These exceptions indicate that the density composition relationship is valid only for a particular class of DNA, possibly for native two-stranded DNA containing no unusual bases.

<sup>5</sup> It might be of interest to compare our estimate of the upper bound on the standard deviation  $\sigma_{GC}$  with the value to be expected for a population of DNA molecules each containing N base pairs assembled at random from a collection of base pairs in which guanine-cytosine occurs with fixed probability one half.

The standard deviation  $\sigma_{GC}$  for this case is just  $\frac{1}{2\sqrt{N}}$ . Taking  $10^4$  as a reasonable value of N

for bacterial DNA<sup>1</sup> there results the value  $\sigma_{GG} = 5 \times 10^{-3}$ , well below the upper limit set by the density gradient experiments. Accordingly, such statistical heterogeneity is not ruled out.

## II

THE MOLECULAR ARRANGEMENT OF THE CONSERVED  
 SUBUNITS OF THE DESOXYRIBONUCLEIC ACID  
 OF ESCHERICHIA COLI

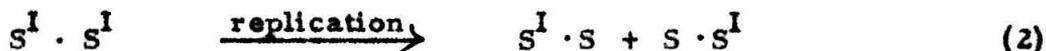
## A. Introduction

1. Statement of the Problem

The DNA molecules of E. coli consist of a pair of submolecular structures or "subunits," which may be represented



During biological replication, the subunits of a DNA molecule are segregated, one into each of two newly formed molecules (1). When isotopically labelled E. coli are transferred to normal culture medium, newly formed subunits have the isotopic composition of the new culture medium, and the replication of the labelled molecules may be written

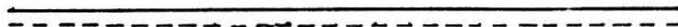


Molecules in which one subunit is isotopically labelled are called "hybrid." Their buoyant density is intermediate between that of unlabelled and fully labelled molecules if heavy isotopes are used for

labelling. If fragments of hybrid molecules contain equal contributions from labelled and unlabelled subunits, they are called "hybrid fragments."

DNA molecules are highly asymmetric, and there are two extreme possibilities for the arrangement of the subunits within them:

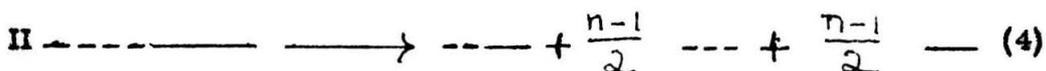
- I The subunits are "side-to-side." Each subunit extends the full length of the molecule.



- II The subunits are "end-to-end." Each extends only half the length of the molecule.



To distinguish between these two possibilities, it is sufficient to break hybrid molecules into segments and examine their density-gradient distribution. Only if the subunits are side-to-side will all of the fragments themselves be hybrid.



## 2. The Initial Approach to the Problem

To calculate the density-gradient distribution of fragments of hybrid DNA from each of the two models for the arrangement of the subunits, it is necessary to know the joint distribution of the fragments with respect to molecular weight and density in each of the two cases. By making certain reasonable assumptions at the outset, we hoped to approximate the correct density-gradient distribution.

We assumed that each molecule is susceptible to fragmentation at a limited number  $N$  of points along its length. Rupture is random and involves no change in density, in the sense that the volume-average density of the two fragments produced is identical to the density of the original molecule. The density of a molecule is a linear function of its isotopic composition and of its nucleotide composition, and these two parameters are mutually independent. The distribution of nucleotides among the molecules has been taken as uniform. This assumption is essentially equivalent to the assumption that nucleotide pairs are distributed randomly, with a fixed probability of occurrence of the guanine-cytosine pair given by the mole fraction of guanine cytosine (GC) in the sample. Using these assumptions, the equilibrium distribution of hybrid DNA at various stages of fragmentation was calculated for each model (Appendix 1, fig. 1).

The family of distributions calculated for each model was compared with an observed family of distributions generated by progressive fragmentation of a sample of hybrid DNA (figs. 2,3). While the observed distributions resembled distributions calculated from the side-to-side model for the arrangement of the subunits, they differed substantially from those calculated from the end-to-end model, providing strong support for the conclusion that the subunits are not end-to-end. Subsequently, it was possible to confirm this initial conclusion using an alternative approach.

### 3. A Second Approach to the Problem

By considering the total variance\* of the density-gradient distribution  $\bar{\sigma}_T^2$  rather than the distribution itself, we can separate, without loss of generality, the effects of isotopic heterogeneity from those of other factors which determine the density-gradient distribution. It can be shown (Appendix 2) that the variance may be written

$$\bar{\sigma}_T^2 = \bar{\sigma}_I^2 + \bar{\sigma}_{M_n}^2 + \bar{\sigma}_X^2 \quad g^2 \text{ cm}^{-6} \quad (5)$$

---

\*The total variance in  $g^2 \text{ cm}^{-6} (\bar{\sigma}_T^2)$  is related to the variance in  $\text{cm}^2 (\sigma^2)$  by the expression

$$\bar{\sigma}_T^2 = \left( \frac{d\rho}{dr} \right)^2 \sigma^2 \quad g^2 \text{ cm}^6 \quad \text{where } \frac{d\rho}{dr} \text{ is the density gradient } g \text{ cm}^{-4}$$

It is independent of the speed of the rotor.

where  $\bar{\sigma}_{M_n}^2$ ,  $\bar{\sigma}_I^2$  are the contributions of thermal motion and isotopic inhomogeneity to the variance, and  $\bar{\sigma}_X^2$  is the contribution of non-isotopic density heterogeneity. The contribution of non-isotopic density heterogeneity may be equal to the contribution of heterogeneity in nucleotide composition among the molecules  $\bar{\sigma}_{GC}^2$  (2,3), or may include the effects of DNA denaturation, or of protein impurities in the DNA. The only assumption involved in formula 5 is that the isotopic composition of a fragment of hybrid DNA is not correlated with any other cause of density heterogeneity. In the subsequent discussion, it will be convenient to assume that  $\bar{\sigma}_X^2 = \bar{\sigma}_{GC}^2$  except where otherwise noted. The behavior of  $\bar{\sigma}_{M_n}^2$  and  $\bar{\sigma}_{GC}^2$  as functions of  $k$ , the average number of scissions per molecule, is the same for unlabelled DNA, hybrid "end-to-end" DNA, and hybrid "side-to-side" DNA. These three "kinds" of DNA differ with respect to the behavior of  $\bar{\sigma}_I^2$ .

For unlabelled DNA obviously  $\bar{\sigma}_I^2 = 0$ . Since in this case  $\bar{\sigma}_{GC}^2 = \bar{\sigma}_T^2 - \bar{\sigma}_{M_n}^2$  and  $\bar{\sigma}_{M_n}^2$  may be obtained from sedimentation data, the behavior of  $\bar{\sigma}_{GC}^2$  may be determined by an experimental investigation of unlabelled DNA.

For "side-to-side" hybrid DNA,  $\bar{\sigma}_I^2 = 0$  also, since the fragments are isotopically homogeneous.

For "end-to-end" hybrid DNA,  $\bar{\sigma}_I^2 > 0$  for  $k > 0$ . In this case  $\bar{\sigma}_I^2$  is a monotonic increasing function of  $k$ , bounded above by

the square of the density separation of the hybrid and unlabelled DNA bands. Thus we may write

$$(\rho_H - \rho_U)^2 \geq \bar{\sigma}_I^2(k) > 0 \text{ g}^2 \text{ cm}^{-6} \quad \text{for } k > 0 \quad (6)$$

where  $\rho_H$ ,  $\rho_U$  are the mean densities of hybrid DNA and unlabelled DNA. The rate with which  $\bar{\sigma}_I^2$  approaches its upper bound depends entirely on the manner in which the molecules are fragmented. If fragmentation begins by introducing a single break into the center of each molecule, then  $\bar{\sigma}_I^2$  would increase linearly from  $\bar{\sigma}_I^2(0) = 0$  to  $\bar{\sigma}_I^2(1) = (\rho_H - \rho_U)^2$  more rapidly than in any other case. In the other extreme case, breaks would occur only at the ends of molecules, releasing fragments so small that  $\bar{\sigma}_I^2(k)$  would increase as slowly as one could desire. Under these circumstances,  $M_w$ , the weight-average molecular weight of the sample, would fall much more slowly than  $M_n$ , the number-average molecular weight, and the ratio  $M_w/M_n$  would increase indefinitely. A third possibility, intermediate between the two extremes discussed above, is random fragmentation, in which the  $M_w/M_n$  ratio rapidly approaches 2. In this case, the behavior of  $\bar{\sigma}_I^2(k)$  can be determined from previously calculated distributions for the end-to-end model, using the expression

$$\frac{\bar{\sigma}_I^2(k)}{(\rho_H - \rho_U)^2} \text{ (calculated)} = \left[ \frac{\bar{\sigma}_T^2(k)}{\bar{\sigma}_{M_n}^2(0)} \text{ (calc)} - \frac{\bar{\sigma}_{M_n}^2(k)}{\bar{\sigma}_{M_n}^2(0)} \right] \frac{\bar{\sigma}_{M_n}^2(0)}{(\rho_H - \rho_U)^2} \text{ (calc)} \quad (7)$$

For  $k = 0.5$  and  $k = 2$ ,  $\frac{\bar{\sigma}_T^2(k)}{\bar{\sigma}_{M_n}^2(0)} \text{ (calc)}$  may be obtained by numerical integration, using the calculated distributions of figure 1,\*  $\frac{\bar{\sigma}_{M_n}^2(k)}{\bar{\sigma}_{M_n}^2(0)}$

is given by the relationship

$$\frac{\bar{\sigma}_{M_n}^2(k)}{\bar{\sigma}_{M_n}^2(0)} = k + 1 \quad (8)$$

Thus, if fragmentation is random, the determination of  $\rho_H - \rho_U$  for a given experiment fixes the behavior of  $\bar{\sigma}_I^2(k)$  expected for the end-to-end model.

#### 4. Outline of Experiments

Sonicated unlabelled DNA was studied to determine the manner in which the DNA had been fragmented, and thus clearly define the expected behavior of  $\bar{\sigma}_I^2(k)$ . The dependence of the ratio  $M_w/M_n$  on  $k$  was estimated from sedimentation data in the following manner. Using an empirical relationship (4) (fig. 5a),  $M_w$  was obtained from

the infinite dilution value of the sedimentation coefficient  $S_0$  and

\* For these distributions we have  $\frac{\bar{\sigma}_{M_n}^2(0)}{(\rho_H - \rho_U)^2} \text{ (calc)} = \frac{1}{64}$

$M_w/M_n$  was estimated from distributions of sedimentation coefficients

(5). Values of  $k$  were obtained from the relationship

$$\frac{M_n(0)}{M_n(k)} = k + 1$$

The behavior of  $\bar{\sigma}_{GC}^{-2}$  as a function of  $k$  was evaluated from the expression

$$\bar{\sigma}_{GC}^{-2} = \bar{\sigma}_T^{-2} - \bar{\sigma}_{M_n}^{-2} \quad g^2 \text{ cm}^{-6} \quad (9)$$

For a DNA sample,  $\bar{\sigma}_{M_n}^{-2}$  is obtained from the formula (6)

$$\bar{\sigma}_{M_n}^{-2} / \left( \frac{d\rho}{dr} \right)_r^2 = RT \left( \bar{v} \left( \frac{d\rho}{dr} \right)_r \omega^2 r M_n \right)^{-1} g^2 \text{ cm}^{-6} \quad (10)$$

where  $\bar{v}$ ,  $\left( \frac{d\rho}{dr} \right)_r$ ,  $\omega$ ,  $r$ ,  $T$ ,  $R$  are the partial specific volume of the DNA in CsCl solution, the value of the density gradient at  $r$ , the angular velocity of the rotor, the distance from the rotor center, the absolute temperature and the gas constant.

Having established the behavior of  $\bar{\sigma}_{M_n}^{-2}(k)$  and  $\bar{\sigma}_{GC}^{-2}(k)$ , and defined the behavior of  $\bar{\sigma}_T^{-2}(k)$  for each of the two models for the arrangement of the subunits, it was possible to predict the expected behavior

of  $\bar{\sigma}_T^2(k)$  in each of the two cases. Experiments were carried out to compare the observed value of  $\bar{\sigma}_T^2$  for fragments of hybrid DNA with the values predicted in this way. In addition, the experiments were designed so that the distribution of the fragments of hybrid DNA could be compared in detail with the distribution expected for the side-to-side model as is explained below.

If the side-to-side model for the arrangement of the subunits is correct, density-gradient distributions of equally fragmented unlabelled and hybrid DNA should differ only in their mean densities. If the end-to-end model is correct, they should differ additionally in a manner that satisfies the relationship

$$\bar{\sigma}_T^2(\text{hybrid}) - \bar{\sigma}_T^2(\text{unlabelled}) = \bar{\sigma}_I^2(k) \text{ g}^2 \text{ cm}^{-6} \quad (11)$$

In order to compare the density-gradient distributions of unlabelled and hybrid DNA at an equal degree of fragmentation, it is sufficient to isolate, fragment and band the two kinds of DNA as a mixture. It is thus possible to determine in detail the manner in which the density-gradient distribution of fragmented hybrid DNA differs from the distribution expected for the side-to-side model. Since the two distributions overlap to some degree, they cannot be compared in their entirety. If the degree of overlap is not too great, it is possible to compare the "dense" half of the hybrid DNA band (H(+)) and the "light" half of the

unlabelled band (U(-)). For symmetrical bands, this is a valid procedure, and the band variance  $\bar{\sigma}_T^2$  is equal to the second moment of each half band (-)(+) around the band mode.

$$m_2(-) = m_2(+) = \bar{\sigma}_T^2 \quad g^2 \text{ cm}^{-6} \quad (12)$$

Differences between  $m_2(-)$ ,  $m_2(+)$  for unlabelled DNA at various stages of fragmentation were investigated and experimental conditions in which a comparison of half bands is valid were established. A detailed comparison of distributions for unlabelled and hybrid DNA fragments indicated that the end-to-end model was not correct.

## B. Experimental Part

### 1. Isolation of DNA

DNA was isolated from exponentially growing cultures of E. coli strain K12, according to a detergent method described by Marmur (7). In one experiment (figs. 2 and 3), E. coli strain B was used, and hybrid DNA was isolated by preparative density gradient centrifugation of a sodium dodecylsulfate lysate.

### 2. Transfer Experiments

For preparation of hybrid DNA, bacteria were grown at 30° C. in a <sup>13</sup>C and <sup>15</sup>N containing algal hydrolysate medium (8), diluted to provide a saturation titer of  $2 \times 10^8$ . Titers were determined by counting the bacteria in a Petroff Hauser chamber. An equal volume of warmed, unlabelled culture medium, capable of providing a saturation titer of  $2 \times 10^9$ , was added to the culture near the end of the exponential phase of growth. The unlabelled medium contained casamino acids, salts, ammonium chloride, and 10 µg/ml of the ribosides of guanine, cytosine, adenine, and uracil. After transfer to the unlabelled medium, the bacteria were allowed two generations of growth, then were chilled and harvested.

### 3. Fragmentation of DNA

DNA in saline citrate buffer (0.015 M sodium citrate, 0.15 M sodium chloride) was fragmented by treatment in a 9 kilocycle sonic oscillator (Raytheon Co., Manchester, New Hampshire). For sonic irradiation, 0.2 to 5 ml. of DNA solution at a concentration of 100 to 400  $\mu\text{g}/\text{ml}$  was placed in a snap-top polyethylene vial, which was floated in water in the cup of the sonic oscillator. Nitrogen was bubbled through solutions just before irradiation, except in one experiment, in which the free radical scavenger AET (amino ethyl isothiuronium bromide hydrobromide, Schwartz Laboratories, Mount Vernon, N.Y.) was used. After passage through a dialysis membrane to remove non-dialysable material, a solution of AET was brought to pH 7.5 with NaOH, and added to the DNA solution to give a final concentration of 5%. After irradiation, the sonicate was dialysed against saline citrate to remove AET. In one experiment, fragmentation was accomplished by vigorous manual shaking of DNA in CsCl solution (figs. 2 and 3).

### 4. Electron Microscopy

Preparation of samples for electron microscopy was carried out by the technique of Hall (9), modified in that a carbon substrate was used instead of the SiO and collodion substrates. Micrographs were made on Kodak lantern slides with an RCA 2A electron microscope.\*

---

\*Kindly made available by Mr. E. Bowler, Electron Microscopy Section, Dept. of Engineering, UCLA.

## 5. Ultracentrifuge Studies

### Photography

For sedimentation velocity and density gradient studies, a Spinco Model E ultracentrifuge and standard ultraviolet absorption optics were used. Absorption photographs of the cell contents were made on Kodak commercial film, and developed 5 minutes in Kodak D11 developer at 20° C. They were converted to tracings of relative absorption versus distance from the rotor center, using a double beam recording microdensitometer (Joyce Loebel, Ltd., Newcastle on Tyne, England) with a slit width corresponding to  $0.3$  mm. in the cell. Linearity of film response was judged by means of a reference aperture (10) within the rotor, and for equilibrium runs, by means of the following additional criterion. A series of exposures of geometrically increasing duration was made. Since maximum contrast is attained when the film response is linear, under and overexposures could be identified by a relative diminution of the band height.

For all equilibrium runs, and most of the velocity runs, the standard ultraviolet optics were supplemented with a 2 mm. thick filter (Corning #9863) placed above the light source to eliminate visible light. With a pyrex filter placed above the #9863 filter, to check for transmission of visible light, exposures of one hour resulted in no visible film blackening.

### Sedimentation Velocity Studies

Velocity runs were carried out at 20° C. in saline citrate buffer at rotor speeds between 23,150 RPM and 44,770 RPM. At the DNA concentrations employed, there was no speed effect (11) on the sedimentation coefficients of our preparations. Optical densities ( $OD_{260}$ ) of DNA solutions were measured in a Beckman DU spectrophotometer. Depending on the optical density of the solution, one of the following centrifuge cell centerpieces was used:

- a) epon, 3 mm. optical path, 4°
- b) anodized aluminum, 12 mm. optical path, 4°
- c) epon, 30 mm. optical path, 2°

Sedimentation coefficients were calculated from the 50% points of boundary tracings (12), with the usual viscosity and density corrections. Samples were studied at several concentrations, and many duplicate runs were performed, both by the resuspension of sedimented DNA and by the use of aliquots of a common DNA solution. The concentration of DNA in each run was estimated from the sedimenting fraction of ultraviolet absorbing material and the total optical density of the solution, assuming that a DNA concentration of 50  $\mu\text{g}/\text{ml}$  corresponds to an optical density of 1.0 (4).

In contrast to results obtained with higher molecular weight bacteriophage DNA (13), passing our preparations through hypodermic

needles (#23 needle, 0.5 ml. per minute) did not reduce the sedimentation coefficient. However, to minimize the shearing forces involved in filling analytical cells, 1 ml. pipettes with orifices at least 2 mm. in diameter were used instead of hypodermic needles. DNA solutions were slowly pipetted into partially assembled cells, open at one end.

### Density Gradient Studies

Density gradient centrifugations were carried out with optical grade CsCl (Maywood Chemical Company, Maywood, New Jersey). Densities of CsCl solutions were obtained using the relationship (14)

$$\rho_{25^{\circ} \text{C.}} = 10.8601 \eta_D^{25^{\circ} \text{C.}} - 13.4974 \pm 0.002 \text{ g cm}^{-3} \quad (13)$$

Refractive index measurements were made with a Zeiss refractometer. CsCl solutions of density  $\rho_{25^{\circ} \text{C.}} = 1.71 \text{ g cm}^{-3}$  containing 0.5 to 5  $\mu\text{g}$  DNA were buffered at pH 8.5 with 0.002 M tris(hydroxymethyl)amino-ethanol. They were centrifuged at 44,770 RPM or 35,600 RPM in an analytical cell with a Kel F centerpiece, 12 mm. optical path, and 1° negative wedge window. A calculation was made of the maximum time necessary for attainment of equilibrium after formation of the density gradient. The following expression<sup>ion</sup> was used (Appendix 3).

$$t_{\text{DNA}}^* < \frac{10^9 (1.2 - \log \sigma^2)}{S \omega^4 r^2} \quad \text{sec} \quad (14)$$

where  $t_{\text{DNA}}^*$  is the time at which the variance of a density gradient distribution of DNA is within 1% of its equilibrium value. Attainment of equilibrium was verified by showing that tracings of photographs taken at 12 to 24 hour intervals were superimposable. For each equilibrium centrifugation (Figs. 8-11, 14, 16) the calculated upper bound on  $t^*$  was in good agreement with the time at which equilibrium was attained, as judged by the above mentioned criterion.

## C. Results

### 1. Properties of Sonic Fragments of Unlabelled DNA

DNA fragments (samples II, III, and IV) were prepared by sonication of purified E. coli K12 W677\* DNA (sample I). A summary of data regarding each sample may be found in table 1. The free radical scavenger AET was used during the preparation of sample IV. In the case of samples II and III, nitrogen was bubbled through the solution just before sonication.

The effect of sonication on the lengths of DNA molecules is demonstrated in electron micrographs of samples I and IV (fig. 4a, b). The spraying procedure involved in sample preparation is known to produce molecular scissions and few molecules of length greater than 2 microns survive (15). However, the fragments in sample IV were considerably shorter than those in sample I, presumably as a result of the sonic treatment.

The effect of sonication on the molecular weight of the DNA was determined from the infinite dilution value of the sedimentation coefficient  $S_{20,w}^{50}$ , using the empirical relationship shown in figure 5a. The extrapolation of sedimentation data to infinite dilution was carried out as shown in figure 6, and the values of  $S_0$ , the extrapolated sedimentation coefficient, and  $M_w$  for each sample may be found in table 1. The possibility that the predominant products of sonication of DNA were

---

\*Paris strain, kindly provided by Dr. G. Bertani.

not macromolecular fragments, and that the sedimentation data was not representative of the bulk of the sample was considered. It was experimentally verified that sonication for the durations employed in these experiments:

- a) had no measurable effect on the optical density ( $OD_{260}$ ) of DNA solutions,
- b) did not increase the proportion of non-sedimenting, ultra-violet absorbing material in samples.

In order to investigate the manner in which the molecules had been fragmented, integral distributions of sedimentation coefficients were obtained from sedimenting boundary profiles, and are shown in figure 7. Values of  $M_w/M_n$  were calculated from them, and are presented in table 1 along with values of  $k$  obtained using formula 8.

For sample II, the  $M_w/M_n$  ratio is consistent with the hypothesis that the molecules have a "most probable" distribution of sizes, resulting from random fragmentation. In contrast to results obtained employing a somewhat different sonication procedure (16), this distribution did not persist with increasing  $k$ , the polydispersity in molecular size decreasing in more highly sonicated samples.\* The data suggest that

---

\*Shear degradation of calf thymus DNA by means of a vaporizer operated at a fixed air pressure has been reported to reduce the polydispersity of samples in a similar manner (17). The minimum attainable molecular weight in this case was  $10^6$ . Larger submolecular structures, around  $2.5 \times 10^6$  in molecular weight, have been reported to result from treatment of E. coli B (18) DNA with chymotrypsin, ~~with heat~~, or by shaking DNA solutions with chloroform octanol.

$7 \times 10^5$  represents a lower limit both for the size of fragments which are broken from larger molecules and for the size of fragments which are themselves broken by sonic vibration. The attainable molecular weight reduction has been found to depend on the precise conditions of sonication, including such factors as the position and volume of the sample, the composition and geometry of the sample container, and the condition of the sonic oscillator.

### The Calculation of $\bar{\sigma}_I^{-2}(k)$ for the End-to-End Model

From the behavior of the ratio  $M_w/M_n$  as a function of  $k$ , we have concluded that fragmentation was essentially random from  $k = 0$  to  $k = 3.4$ . In this range, it is valid to use the calculated distributions of figure 1 to predict the behavior of  $\bar{\sigma}_I^{-2}(k)$ . Values of  $\frac{\bar{\sigma}_I^{-2}(k)}{(\rho_H - \rho_U)^2}$  (calc) for  $k = 0.5$  and  $k = 2$  were obtained from the distributions of figure 1 using formula 7. They were converted to values of  $\frac{\bar{\sigma}_I(k)}{\bar{\sigma}_{M_n}(0)}$  (calc), and plotted in figure 17, using the ratio  $\frac{\rho_H - \rho_U}{\bar{\sigma}_{M_n}(0)} = \frac{19}{2}$  (0) obtained from a transfer experiment which will be described. A smooth curve was drawn through the two points, asymptotic to the line  $\frac{\bar{\sigma}(k)}{\bar{\sigma}_{M_n}(0)} = \rho_H - \rho_U$ . This curve has been taken to represent the dependence of  $\bar{\sigma}_I^{-2}$  on  $k$  expected for random fragmentation of the DNA. It provides an estimate of the fraction of  $\rho_H - \rho_U$  attained by  $\bar{\sigma}_I$  as a function of  $k$ , independent

of the values of  $\rho_H - \rho_U$  and of  $\bar{\sigma}_{M_n}^{-2}(0)$ . Using this curve, we conclude that  $\sigma_I^2(k)$  should attain 70% of its maximum by  $k = 3.4$ . For  $k > 5$ ,

$\sigma_I^2(k)$  should increase as rapidly as in the case of random fragmentation, since few fragments of molecular weight less than  $7 \times 10^5$  are produced. Thus the behavior of  $\bar{\sigma}_I^{-2}(k)$  expected on the basis of the end-to-end model is adequately represented by the curve of figure 17.

The Calculation of  $\bar{\sigma}_{M_n}^{-2}(k)$  from Sedimentation  
Velocity Data

For DNA in CsCl solution of density  $\rho = 1.71$  at  $25^\circ \text{C}$ . and for  $r = 6.5$ , we may write (6)

$$\left( \frac{d\rho}{dr} \right)_r = 8.35 \times 10^{10} w^2 r \text{ g cm}^{-4} \quad (15)$$

Thus equation 10 becomes

$$\bar{\sigma}_{M_n}^{-2} = \frac{35.5}{M_n} \text{ g}^2 \text{ cm}^{-6} \quad (16)$$

Having obtained  $M_n$  from sedimentation data as outlined earlier, we calculated  $\bar{\sigma}_{M_n}^{-2}$ . Values of  $\bar{\sigma}_{M_n}^{-2}$  for samples I-IV may be found in table 1. The behavior of  $\bar{\sigma}_{M_n}^{-2}(k)$  is represented in figure 17 by a plot of  $\frac{\bar{\sigma}_{M_n}^{-2}(k)}{\bar{\sigma}_{M_n}^{-2}(0)} = (k+1)^{1/2}$  versus  $k$ .

### The Calculation of $\bar{\sigma}_{GC}^2(k)$ from Density-Gradient Data

Using equation 9, and evaluating  $\bar{\sigma}_T^2(k)$  numerically from the observed density-gradient distribution, values of  $\bar{\sigma}_{GC}^2(k)$  were calculated. Values of  $\bar{\sigma}_{GC}^2$  and  $\bar{\sigma}_T^2$  for samples I-IV may be found in table 1. A plot of  $\frac{\bar{\sigma}_{GC}^2(k)}{\bar{\sigma}_{M_n}^2(0)}$  versus  $k$  was constructed as follows: Values of  $\bar{\sigma}_{GC}^2$  for four different values of  $k$  were obtained from data for samples I-IV, divided by  $\bar{\sigma}_{M_n}^2(0)$  (from data for sample I), and plotted in figure 17. A smooth curve connecting the points represents the behavior of  $\bar{\sigma}_{GC}^2(k)$ .

### The Calculation of $\bar{\sigma}_T^2(k)$ Expected for the Two Models for the Arrangement of the Subunits

Values of  $\bar{\sigma}_T^2(k)$  expected for each model for the arrangement of the subunits were calculated from figure 17, using equation 5 and values of  $\rho_H - \rho_V$  and  $\bar{\sigma}_{M_n}(0)$  obtained in a subsequent experiment. Figure 18 shows the behavior of  $\bar{\sigma}_T^2$  as a function of  $k$ , expected on the basis of each of the two models.

### The Determination of $\bar{\sigma}_T^2$ from Half Bands

With the exception of one sample in which a small amount of DNA had apparently been denatured, bands of sonicated DNA did not show significant skewness, and the following relationship was experimentally verified (Appendix 5).

$$\bar{\sigma}_T^2 \approx m_2(-) \approx m_2(+)$$
(17)

where  $m_2(-)$ ,  $m_2(+)$  are the second moments of the half bands (-) (+) around the band mode. Because of band symmetry, it is possible to evaluate  $\bar{\sigma}_T^2$  for partially overlapping bands of sonicated unlabelled and hybrid DNA by evaluating  $m_2(U(-))$  and  $m_2(H(+))$ .\*

## 2. Properties of Sonic Fragments of Transfer DNA

DNA was isolated from E. coli K12 two generations after transfer from labelled to unlabelled culture medium. The method of Marmur (7) was used for isolation. In the density gradient, this "transfer" DNA (sample V) formed two slightly overlapping bands (fig. 14), one at a density expected for unlabelled DNA, and the other at a density expected for hybrid  $^{13}\text{C}$ ,  $^{15}\text{N}$  labelled DNA. Sedimentation velocity studies (fig. 12) indicated that the transfer DNA had a weight average molecular weight of  $12 \times 10^6$ . The concentration dependence of  $S_{20,w}^{50}$  and the distribution of sedimentation coefficients (fig.13) were similar to those of unlabelled DNA previously isolated.

---

\*In case of DNA denaturation, the assumption of band symmetry for the sonicated hybrid DNA band is incorrect and we may write

$$m_2(H(+)) > \bar{\sigma}_T^2 \text{ g}^2 \text{ cm}^{-6}$$

Thus the assumption of symmetry would lead to an overestimation of  $\bar{\sigma}_T^2$  favoring the end-to-end model.

Because of slight overlap of the two bands of sample V, the (-) half of the unlabelled DNA band (U(-)), and the (+) half of the hybrid band (H(+)) were studied. Values of various parameters for the two half bands may be found in table 2. The fact that  $m_2(U(-)) > m_2(H(+))$  was consistent with the mild band skewness previously observed for unlabelled DNA before sonication (Appendix 5). The magnitudes of  $m_2(H(+))$ ,  $M_w/M_n$ , were also comparable to those for sample I.

Two additional samples of transfer DNA (VI and VII) were prepared by sonication of aliquots of sample V in the absence of AET. Samples VI and VII were studied in the CsCl density gradient (figs. 15 and 16) and sedimentation velocity analysis was carried out on sample VII (figs. 12 and 13). The observed values of the physical parameters may be found in table 2.

The effects of sonication on the sedimentation behavior of the transfer DNA were similar to its effects on unlabelled DNA. Both unlabelled and transfer DNA were reduced to the same range of molecular size by sonication, and in this range, the size polydispersity as judged from distributions of sedimentation coefficients (figs. 6 and 13) was comparable for the two kinds of DNA. As the transfer DNA was smaller in size before sonication, fewer scissions per molecule were attained in this case ( $k = 7$ ).

Predictions for the End-to-End Model

If the end-to-end model is correct, at  $k = 7$  more than 80% of the original mass of the DNA would no longer have the mean density of hybrid DNA  $\rho_H$ , but would consist of unlabelled and fully labelled fragments with mean densities  $\rho_U, \rho_L$ . Sonication of the transfer DNA (sample V) should increase the ratio of unlabelled to labelled DNA by a factor of two, resulting in a similar change in the ratios of heights and areas of the unlabelled and labelled DNA bands.

Predictions for the Side-to-Side Model

If the side-to-side model is correct, the hybrid and unlabelled sonicated DNA s should have the same density-gradient distribution, but with different means. If the unlabelled DNA band is symmetrical, there will be a position in the cell where the CsCl density is  $\rho_U^*$  and such that

$$\rho_U - \rho_U^* = \rho_H - \rho_U \quad (18)$$

and

$$C_U(\rho_U) = C_U(\rho_U^*) \quad (19)$$

where  $C_U(\rho_U), C_U(\rho_U^*)$  are the concentrations of fragments of unlabelled DNA for  $\rho_U, \rho_U^*$ . Thus if the fragments of hybrid DNA and unlabelled DNA are banded together, and if  $C_H(\rho_U) \ll C_U(\rho_U)$   $C_U(\rho_H) \ll C_H(\rho_H)$  then for the mixture we may write

$$\frac{C(\rho_U^*)}{C(\rho_U)} = \frac{C(\rho_L)}{C(\rho_H)} \quad (20)$$

In the side-to-side case, the ratio of unlabelled to labelled DNA in the unsonicated sample of transfer DNA should not be affected by sonication. Relative heights and relative areas of the bands should remain essentially constant.

#### Characteristics of the Experimental Distributions

Having presented detailed predictions of the equilibrium distribution of the sonicated transfer DNA expected for each of the two models for the arrangements of the subunits, we now consider the experimental distributions. Since we have sedimentation velocity data for sample VII only, our discussion refers primarily to the density gradient distribution for this sample (fig. 16). The distribution for sample VI at a lower  $k$  value, is completely consistent with that of sample VII (fig. 15).

The observed density gradient distribution for sample VII is representative of the entire sample, since essentially all of the DNA introduced into the analytical cell can be accounted for in the band, with the aid of the reference aperture. Referring to table 2, we note that the density difference between the modes of the unlabelled and hybrid DNA bands decreased by no more than  $0.001 \text{ g cm}^{-3}$  after

sonication. This effect was attributed to the overlapping of the two bands, which contributes to the concentration of the DNA preferentially between the band modes. The convergence of the two band modes increased the observed values of  $m_2(U(-))$  and  $m_2(H(+))$  by less than 5%.

The shapes of the half bands  $U(-)$  and  $H(+)$  are highly similar, and  $m_2(U(-))$  and  $m_2(H(+))$  agree well with each other. Values of  $m_2(H(+))$  for samples VI and VII are compared with curves of  $\sigma_T^{-2}(k)$  (calculated, side-to-side) and  $\sigma_T^{-2}(k)$  (calculated, end-to-end) in figure 18. The value of  $k$  for sample VII was determined by sedimentation studies as described, and the value of  $k$  for sample VI was estimated from the value of  $m_2(U(-))$ . The experimental values clearly indicate that the end-to-end model is not correct. If the determination of  $\sigma_T^{-2}$  or  $k$  were in error by 400%, the experimental result would have been intermediate between the two models, and the range of error is certainly less than 20%. The height and area of  $H(+)$  relative to those of  $U(-)$  were essentially unaffected by sonication. The small decrease in area of  $U(-)$  probably reflects a decreased skewness of the band of unlabelled DNA as a result of sonication, as was found for unlabelled DNA as described in Appendix 5. The concentration of DNA at  $\rho_L$  is less than 3% of that at  $\rho_H$ , and is essentially the same as the DNA concentration at  $\rho_U^*$ , relative to that at  $\rho_U$ .

$$\frac{C(\rho_U^*)}{C(\rho_U)} \sim \frac{C(\rho_L)}{C(\rho_H)} < 0.03 \quad (21)$$

Thus no evidence of fully labelled fragments could be detected, and it was possible to set an upper limit of 3% for the relative concentration of fully labelled fragments resulting from the sonication of the hybrid DNA.

D. Summary

It has been demonstrated that sonic fragments of hybrid DNA preserve their hybrid character and that the subunits of DNA are not in an end-to-end arrangement. Knowledge of the mass per unit length of these fragments would fix an upper limit for the number of polynucleotide strands within a single subunit.

## APPENDIX :

Calculation of equilibrium distributions for fragments of hybrid DNA from models for the arrangement of the subunits.

The most general expression which may be written for the density gradient distribution of fragments of hybrid DNA is

$$C(\rho) = \int \int \int_{M \ I \ GC} C(I, GC, M) \exp\left(-\frac{1}{2} \frac{[\rho - \rho_0(I, GC)]^2}{\sigma_M^2}\right) dGC \ dI \ dM \quad (1)$$

where  $C(\rho)$  is the DNA concentration at density  $\rho$ ,  $\sigma_M^2$  is the variance ( $\text{g}^2 \text{cm}^{-6}$ ) for bands of DNA of molecular weight  $M$  and density  $\rho_0$ , and  $C(I, GC, M)$  is the concentration of DNA of molecular weight  $M$ , nucleotide composition  $GC$  and isotope composition  $I$ , in the total sample. To calculate the equilibrium distribution from this expression, it is sufficient to know the distribution of the DNA fragments with respect to the parameters  $GC$ ,  $I$  and  $M$ . We may assume that  $GC$  is independent of  $I$  and  $M$ , and concern ourselves with the distribution of the DNA fragments with respect to nucleotide composition alone. If one assumes that base pairs are distributed randomly among the fragments of DNA, the distribution of guanine cytosine base pairs within each size class is binomial, with variance  $np(1-p)$ . The variance of

the mole fraction of guanine cytosine for DNA is given by

$$\sigma_{GC}^2 = \frac{1}{4n} = \frac{1}{4} \frac{1}{M/M_0} = \frac{220}{M} \quad (2)$$

where  $M$ ,  $M_0$ ,  $n$  are the molecular weight of the DNA, the weight of a nucleotide pair (880 for the  $Cs^+$  salt of DNA), and the number of nucleotide pairs. The assumption of randomness in the distribution of GC pairs is formally equivalent to the assumption that the DNA is homogeneous in nucleotide composition, since equilibrium distributions in the two cases are nearly identical. Consider an equilibrium distribution of molecules of isotopic composition I and molecular weight  $M$  in which GC pairs are randomly distributed. The distribution is gaussian (19) with variance given by\* (Appendix 2)

$$\sigma_T^2 = \sigma_M^2 + \sigma_{GC}^2 = \frac{35.5}{M} + \frac{2.2}{M} = \frac{37.7}{M} \quad (3)$$

and is identical to a distribution for which  $\sigma_{GC}^2 = 0$  if the density scale is multiplied by a factor 0.97. Within a factor of 0.97 in the density distribution, the result is generally valid since any equilibrium distribution is a sum of gaussians.

To complete the calculation of an equilibrium distribution from equation 1 we must consider the distribution of the fragments with respect to their isotopic composition I and molecular weight  $M$ . This

---

\*From this calculation we also note that for molecules in which base pairs are randomly distributed

$$\frac{\sigma_{GC}}{\sigma_{M_n}} = 0.24$$

distribution depends on the nature of the assumptions made regarding the manner in which the DNA is fragmented. We have assumed that breaks occur randomly at a fixed number (N) of possible breaking points along the length of the molecule. The observation that breakage is unlikely for fragments below a certain size is consistent with this assumption. It has been verified by calculation that the result is essentially independent of N for  $K/N < 0.1$ .

The calculation is outlined below, and takes the form of a depolymerization analysis. In the case of the sonic fragmentation of DNA, the smallest fragment which can be produced by sonic irradiation may be taken as the "monomer."

The equilibrium distribution of random fragments of a macromolecule in a density gradient.

Consider a solution of polymer molecules A each composed of monomers  $a_i$ ,  $i = 1, \dots, N+1$  (N odd) joined by links  ${}_i a_{i+1}$  with probability of rupture  $p({}_i \bar{a}_{i+1})$ . For the case of random fragmentation of the polymers, the fraction of polymer molecules which will be broken at the link  ${}_i a_{i+1}$  is independent of i

$$p(A / \overline{{}_i a_{i+1}} \in A) = p({}_i \bar{a}_{i+1}) = p \quad (4)$$

and the fraction which will give rise to fragments  ${}_i A_j = a_i a_{i+1} \dots a_{j-1} a_j$  is

$$\begin{aligned}
 P(A/{}_i A_j \in A) = & \quad (1-p)^N && \text{for unbroken molecules } {}_i A_j = A \\
 & p(1-p)^{j-i} && \text{end fragments } i=1 \text{ or } j=N+1 \\
 & p^2(1-p)^{j-i} && \text{middle fragments } i \neq 1 \text{ and } \\
 & && \quad j \neq N+1
 \end{aligned}
 \tag{5}$$

Among the fragments resulting from random breakage, the proportion which are  ${}_i A_j$  is

$$P({}_i A_j) = \frac{P(A/{}_i A_j \in A)}{\sum_{i < j \leq N+1} P(A/{}_i A_j \in A)} = \frac{P(A/{}_i A_j \in A)}{1+Np}
 \tag{6}$$

For the end-to-end model of DNA let the densities of the monomers  $a_i$  be given by

$$\begin{aligned}
 \rho(a_i) = \rho_U & \quad i = 1, \dots, \frac{N+1}{2} && \text{(unlabelled)} \\
 \rho_L & \quad i = \frac{N+1}{2} + 1, \dots, N+1 && \text{(fully labelled)}
 \end{aligned}
 \tag{7}$$

Let  $[B_{UL}]$  denote the set of fragments with U unlabelled and L labelled monomers

$$[B_{UL}] = \left[ \begin{array}{l} \diagup \\ \diagdown \end{array} \begin{array}{l} j-i = U + L - 1 \\ i = \frac{N+1}{2} + 1 - U \\ j = \frac{N+1}{2} + L \\ i > \frac{N+1}{2} \\ j < \frac{N+1}{2} + 1 \end{array} \begin{array}{l} \text{in any case} \\ \text{and} \\ \dots \text{ if } U > 0 \text{ and } L > 0 \\ \dots \text{ if } L = 0 \\ \dots \text{ if } U = 0 \end{array} \right]
 \tag{8}$$

Since there are  $\frac{N+1}{2} - U$  equally likely different types of internal fragments composed of  $U$  unlabelled monomers only, the fraction of fragments containing  $U$  unlabelled and  $L$  fully labelled monomers is given by the matrix element

$$p(B_{UL}) = \sum_{iA_j \in [B_{UL}]} p(iA_j) =$$

$$\frac{p(1-p)^{U+L-1} + p^2(1-p)^{U+L-1} \left( \frac{N+1}{2} - (U+L) \right)}{1+Np} \quad \text{if } U=0 \text{ or } L=0$$

$$\frac{p(1-p)^{U+L-1}}{1+Np} \quad U \text{ or } L = \frac{N+1}{2}$$

$$\frac{(1-p)^{U+L-1}}{1+Np} \quad U \text{ and } L = \frac{N+1}{2}$$

$$\frac{p^2(1-p)^{U+L-1}}{1+Np} \quad \text{otherwise}$$

(9)

The weight fraction of  $B_{UL}$  is the fraction of monomers  $a_i$  in fragments  $iA_j \in [B_{UL}]$

$$F_w(B_{UL}) = p(a_i \in B_{UL}) = \frac{U+L}{N+1} (1+Np) p(B_{UL}) \quad (10)$$

Fragments  $B_{UL}$  have the equilibrium distribution (6)

$$C_{UL}(r) = \frac{1}{\sqrt{2\pi} \sigma_{UL}} \exp\left(-\frac{1}{2} \frac{(r-r_{UL})^2}{\sigma_{UL}^2}\right) \text{ (normalized)} \quad (11)$$

where  $\sigma_{UL}^2 = \frac{\rho_{UL}/\rho_A}{M_{UL}/M_A} (r_A/r_{UL})^2 \sigma_A^2$  is the variance of the distribution of  $B_{UL}$

$\sigma_A^2 = \frac{RT}{M_A \bar{v}_A \left(\frac{d\rho}{dr}\right)_{r_A} w^2 r_A}$  is the variance of the distribution of polymer A

$\rho_{UL} = \rho(B_{UL}) = \frac{U\rho_U + L\rho_L}{U+L}$  is the density of  $B_{UL}$

$M_{UL} = M(B_{UL}) = \frac{U+L}{N+1} M_A$  is the molecular weight of  $B_{UL}$

and distances of band centers from the center of rotation are

$r_A$  for the polymer A

$r_U = r_A - \frac{1}{2} \frac{\rho_L - \rho_U}{(d\rho/dr)_{r_A}}$  for the unlabelled monomers

$r_L = r_A + \frac{1}{2} \frac{\rho_L - \rho_U}{(d\rho/dr)_{r_A}}$  for the labelled monomers

$r_{UL} = \frac{Ur_U + Lr_L}{U+L}$  for the fragments

\*(Assuming  $\bar{v} = 1/\rho$ )

Thus the equilibrium distribution of the fragments is

$$C(r) = \sum_{U, L \leq \frac{N+1}{2}} F_w(B_{UL}) C_{UL}(r) \quad (\text{normalized}) \quad (12)$$

Setting  $\rho_U = \rho_L$  we obtain the equilibrium distribution of fragments for the side-to-side model of DNA. In this case the fraction of fragments of size  $S$  is

$$p(B_S) = \sum_{U+L=S} p(B_{UL}) =$$

$$\frac{p(1-p)^{S-1} (2+p(N-S))}{1+Np} \quad \text{for } S \leq N \quad (13)$$

$$\frac{(1-p)^N}{1+Np} \quad S = N+1$$

the weight fraction is

$$F_w(B_S) = \frac{S}{N+1} (1+Np) p(B_S) \quad (14)$$

and the equilibrium distribution

$$C(r) = \sum_{S=1}^{N+1} F_w(B_S) \frac{1}{\sqrt{2\pi} \sigma_S} \exp\left(-\frac{1}{2} \frac{(r-r_A)^2}{\sigma_S^2}\right) \quad (15)$$

(normalized)

where  $\sigma_S^2 = \frac{N+1}{S} \sigma_A^2$

Calculation of equilibrium distributions according to equations 12 and 15 were carried out with a Burroughs 205 computer for various values of  $N$  and  $k = Np$ , assuming  $\rho_L - \rho_U = 16 \sigma_A \left( \frac{d\rho}{dr} \right)$ . The contribution to the equilibrium distribution of fragments smaller than 0.1 of the initial molecular size was ignored. The results are presented below.

$$\text{Values of } \frac{C(\rho)}{\int C(\rho) d\rho} = 0.399 \frac{C(\rho)}{C(\rho_0)_{(k=0)}}$$

are tabulated at intervals of  $X \bar{\sigma}_{M_n}^{(0)}$ , starting at the mean of the distribution.

## SIDE-TO-SIDE MODEL

	I	II-1	II-2	II-3	
N	9	99	99	99	4
X	0.3	0.3	0.3	0.3	0.5
K	0.5	2	4	8	0
1	.359	.284	.222	.151	.399
2	.345	.276	.217	.149	.352
3	.307	.253	.204	.143	.242
4	.252	.220	.185	.135	.130
5	.193	.181	.161	.123	.054
6	.138	.143	.137	.111	.018
7	.093	.109	.112	.097	.004
8	.059	.081	.091	.084	
9	.037	.059	.072	.071	
10	.022	.042	.056	.060	
11	.014	.030	.043	.050	
12	.009	.022	.033	.041	
13		.016	.025	.033	
		.011	.019	.026	
			.015	.021	
				.017	
				.013	
				.010	
$X \int C dp$	.99	.96	.89*	.89*	.99

\*Low values are due to neglect of small fragments.

## END-TO-END MODEL

	I	IIa1	IIa2	IIa3	IIb1	IIb2	III
N	9	19	19	19	19	19	99
X	0.5	0.3	0.3	0.3	1	1	
K	0.5	0.5	1.0	1.5	0.5	2	2
1	.261	.268	.181	.125	.268	.088	.098
2	.233	.257	.175	.121	.172	.068	
3	.169	.228	.158	.111	.056	.043	
4	.102	.187	.133	.097	.022	.034	
5	.056	.143	.107	.081	.016	.031	
6	.033	.103	.083	.067	.014	.032	
7	.023	.072	.063	.055	.015	.036	
8	.019	.050	.049	.046	.016	.044	
9	.017	.035	.040	.040	.018	.049	
10	.016	.027	.034	.036	.016	.045	
11	.015		.031	.034		.032	
12	.015		.028	.032		.018	
13	.015		.027	.031		.008	
14	.016		.025	.030			
15	.017		.025	.029			
16	.018			.029			
17	.019			.029			
18	.019			.029			
19	.017			.029			
20	.015			.030			
21	.012			.032			
22	.009			.033			
XSCP	.99	*	*	*	.96	.97	

\*The complete distribution was not calculated.

## APPENDIX 2

The relationship between the variance of the equilibrium distribution of fragments of hybrid DNA and the variances of the distributions of nucleotides and heavy isotopes among the molecules.

For molecules heterogeneous in density and molecular weight the variance of the equilibrium distribution  $\sigma_T^{-2}$  may be written (5)

$$\sigma_T^{-2} = \sigma_D^{-2} + \sigma_{M_n}^{-2} \quad (1)$$

where  $\sigma_D^{-2}$  is the variance of the density distribution and  $\sigma_{M_n}^{-2}$  is the contribution of the thermal movement of the molecules to the variance of the equilibrium distribution. Assuming that there are no other causes of density heterogeneity, the buoyant density of a DNA molecule is a linear function of its nucleotide and isotopic composition. For the degree of isotopic labelling involved in these experiments the linear relation is

$$\rho_{25^\circ} = 1.658 + 0.1 \text{ GC} + 0.032 \text{ I} \quad (2)$$

where GC and I are mole fractions of guanine cytosine and heavy isotope. To a first approximation, the isotopic and nucleotide compositions of a fragment of hybrid DNA are independent random variables.\* By an elementary theorem of probability theory (20) we may write

$$\sigma_D^2 = \sigma_{GC}^2 + \sigma_I^2 \quad (3)$$

where  $\sigma_{GC}^2$ ,  $\sigma_I^2$  are the contributions of nucleotide and isotopic heterogeneity to the variance of the density distribution. The parameters  $\sigma_{GC}^2$ ,  $\sigma_I^2$  are related to the variances of the distributions of nucleotides and isotopes  $\sigma_{GC}^2$ ,  $\sigma_I^2$  among the molecules as follows

$$\sigma_{GC}^2 = 10^{-2} \times \sigma_{GC}^2 \quad (4)$$

$$\sigma_I^2 = 10^{-3} \times \sigma_I^2 \quad (5)$$

---

\*The ratio of density labelling for the guanine cytosine and adenine thymine pairs is given by their mass increment ratio, assuming no change in volume. The GC nucleotide pair has 19 carbon atoms and 8 nitrogens, the AT pair 20 carbons and 7 nitrogens. For DNA containing 54%  $^{13}\text{C}$  and 99%  $^{15}\text{N}$  the ratio is

$$\frac{\Delta M(\text{GC})}{\Delta M(\text{AT})} = \frac{0.54(19) + 8}{0.54(20) + 7} = \frac{17.8}{17.2} = 1.03$$

## APPENDIX 3

Calculation of the time required to reach equilibrium for density gradient centrifugation of DNA in CsCl.

Consider a polymer homogeneous in molecular weight  $M$  and buoyant density  $\rho_0$  forming a band in the density gradient. The time dependent differential equation characterizing the approach to equilibrium is written (6)

$$\frac{d}{dx} \left( \frac{dC}{dx} + \frac{x C(x)}{\sigma^2} \right) = \frac{1}{D} \frac{dC}{dt} \quad (1)$$

In this equation  $t$  is the time elapsed during the run,  $x$  is the distance in the cell from that position where the density of the medium is  $\rho_0$ .  $C(x)$  is the concentration of the polymer at  $x$ ,  $D$  is the diffusion coefficient for the polymer. The variance of the equilibrium distribution of the polymer  $\sigma^2$  is given by equation 10 (see text, p. 13).

Using equation 1, we may obtain the following integral equation

$$\int_{-\infty}^{\infty} x^n \frac{d}{dx} \left( \frac{dC}{dx} + \frac{x C}{\sigma^2} \right) dx = \int_{-\infty}^{\infty} \frac{x^n}{D} \frac{dC}{dt} dx \quad (2)$$

The left side of 2 may be integrated by parts as follows

$$\int_{-\infty}^{\infty} x^n \frac{d}{dx} \left( \frac{dC}{dx} + \frac{x C}{\sigma^2} \right) dx = x^n \left( \frac{dC}{dx} + \frac{x C}{\sigma^2} \right) \Big|_{-\infty}^{\infty} - n \int_{-\infty}^{\infty} x^{n-1} \left( \frac{dC}{dx} + \frac{x C}{\sigma^2} \right) dx \quad (3)$$

$$-n \int_{-\infty}^{\infty} x^{n-1} \frac{dC}{dx} dx = -n x^{n-1} C \Big|_{-\infty}^{\infty} + n(n-1) \int_{-\infty}^{\infty} x^{n-2} C dx \quad (4)$$

Thus we obtain the relationship

$$x^n \left( \frac{dC}{dx} + \frac{x C}{\sigma^2} \right) - n x^{n-1} C \Big|_{-\infty}^{\infty} = -\frac{n}{\sigma^2} \int_{-\infty}^{\infty} x^n C dx - n(n-1) \int_{-\infty}^{\infty} x^{n-2} C dx + \frac{1}{D} \int_{-\infty}^{\infty} x^n \frac{dC}{dt} dx \quad (5)$$

For bands whose widths are small compared to the radial width of the liquid column in the centrifuge cell, there exists a time  $t_0$  such that for  $t > t_0$  we have

$$\lim_{|x| \rightarrow \infty} C(x, t) = 0, \quad \lim_{|x| \rightarrow \infty} \left( \frac{dC}{dt} \right) = 0$$

Using these boundary conditions, the left side of equation 5 vanishes, and we have a moment relationship first given by R. Feynman (14)

$$\frac{n}{\sigma^2} A_n(t) + n(n-1) A_{n-2}(t) = \frac{\dot{A}_n(t)}{D} \quad (6)$$

where  $A_n(t)$  is the  $n^{\text{th}}$  moment of the polymer distribution at time  $t$  and  $\dot{A}_n(t)$  is its time derivative. By definition  $\sigma^2 = \frac{A_2(\infty)}{A_0}$

hence for  $n = 2$  equation 6 becomes

$$\dot{A}_2(t) = -\frac{2D}{\sigma^2} [A_2(t) - A_2(\infty)] \quad (7)$$

which has the unique solution

$$\ln \frac{A_2(t) - A_2(\infty)}{A_2(0) - A_2(\infty)} = - \frac{2D}{\sigma^2} t \quad (8)$$

Let us define  $t_{DNA}^*$  as the time of centrifugation for DNA such that  $t > t_{DNA}^*$  implies  $\left( \frac{\sigma^2(t)}{\sigma^2(\infty)} - 1 \right) < 0.01$ , where  $\sigma^2(t) = \frac{A_2(t)}{A_0}$ . We use equation 8 to find an upper bound for  $t_{DNA}^*$ . If  $S$  is the infinite dilution value of the sedimentation coefficient with respect to water at  $20^\circ\text{C}$ , we have

$$D = \frac{RTS}{M(1-\bar{v})} \quad \text{and } \rho = 1$$

assuming  $D$  is independent of the medium. Taking  $\frac{1}{\bar{v}} = 1.71 \text{ g cm}^{-3}$  for solvated DNA, and using equations 10 and (15 of the text (pp. 13 and 15), we rewrite equation 8

$$\ln \left[ \frac{\sigma^2(t)}{\sigma^2(\infty)} - 1 \right] - \ln \left[ \frac{\sigma^2(0) - \sigma^2(\infty)}{\sigma^2(\infty)} \right] = -24 \times 10^{-10} S w^4 r^2 t$$

Since for  $t = 0$ ,  $C$  is independent of  $x$ , we have

$$\sigma^2(0) = \frac{\int_0^{0.7} x^2 dx}{\int_0^{0.7} dx} \approx 0.16$$

and thus obtain

$$t_{DNA}^* < \frac{10^9 (1.2 - \log \sigma^2(\infty))}{S w^4 r^2} \text{ sec}$$

## APPENDIX 4

The determination of molecular weight for samples of DNA.

The conclusions drawn in this study depend importantly on measurements of molecular weight for samples of DNA. Since these measurements utilized a calibration of  $S$  versus  $M$  which had been carried out in other laboratories, it seemed desirable to carefully examine the validity of the calibration and of our application of it.

For DNA of  $M < 5 \times 10^6$ , values of  $M$  obtained by light scattering agree well with values obtained from sedimentation and viscosity measurements, using the Mandelkern-Scheraga equation (21)

$$\beta = \frac{S_o [\eta]^{1/3}}{M_w^{2/3}} \frac{1 - \bar{v} \rho}{\eta_o N} \quad (1)$$

where  $[\eta]$ ,  $S_o$ ,  $M_w$ ,  $\bar{v}$  are the intrinsic viscosity, the infinite dilution value of the sedimentation coefficient, the weight average molecular weight, and the partial specific volume of the polymer;  $\eta_o, \rho$  are the viscosity and the density of the solvent and  $\beta$  is an empirical parameter. The magnitude of  $\beta$  has been studied for flexible chain polymers of different molecular weights in different solvents, for proteins and for TMV, and has been found to be a constant for a given polymeric

type, independent of molecular weight\* and of the solvent, with a value between  $2.1$  and  $2.7 \times 10^6$  (22). For DNA of  $M < 5 \times 10^6$ , as determined by light scattering, values of between  $2.4$  and  $2.9 \times 10^6$  have been reported. For larger molecular weight DNA,  $\beta$  increases in magnitude, reaching  $3.6 \times 10^6$  for a sample (16) of  $M = 7.7 \times 10^6$ . Butler et al. (24) have advanced a theoretical argument indicating that light scattering is inapplicable to the determination of molecular weights above  $5 \times 10^6$  for DNA. They assume a  $\beta$  of  $2.6 \times 10^6$  in calculating  $M$  from sedimentation and viscosity measurements, using equation 1. Doty has adopted the procedure of accepting the light scattering data up to  $8 \times 10^6$  and using the curvature of the  $\beta$  versus  $M$  plot to extrapolate  $\beta$  for  $M$  as high as  $16 \times 10^6$  (4). We have accepted the Doty calibration of  $\beta$ , and used his  $M_w$  versus  $S_o$  plot, as it is not likely to lead to an overestimation of the molecular weight of unsonicated DNA, and thus is not likely to lead to an overestimation of  $k$  for sonicated DNA.

Using the calibration discussed above, the determination of molecular weights below 5 million from sedimentation data is a reasonably reliable procedure. However, DNA sonicates made with and without AET differ in certain respects, and it is necessary to discuss the possible

---

\*For one system, polymethyl methacrylate in acetone,  $\beta$  increases from 2.12 to 2.52 as  $M_w$  increases from 30,000 to 14,000,000 (23).

effects of these differences on our conclusions.

There are a number of indications that the use of AET reduces the damage done to DNA by the free radicals produced during sonication. AET abolishes the effect of DNA concentration on the inactivation of genetic markers by sonication, as well as reducing the absolute amount of inactivation (25). From the band shape of sample III (fig. 10), one might conclude that sonication without AET had resulted in the denaturation of a small fraction of the DNA sample. The data of figure 6 suggest that the concentration dependence of  $S_{20,w}^{50}$  differs for sonicates made with and without AET. Only one sonicate of molecular weight greater than  $10^6$  made without AET has been reported. The values of  $S_0$  and  $M_w$  for this sonicate are inconsistent with the empirical relationship shown in figure 5a, and may indicate that the relationship for sonicates made without AET should be described by the dotted line in figure 5b.

However, the investigators (16) responsible for the values of  $M_w$  and  $S_0$  for this sample also obtained a value of  $[\eta]$  which was consistent with that of other DNA samples of the same molecular weight. Substitution of the values of  $[\eta]$ ,  $M$  and  $S_0$  into equation 1 gave a value of  $\beta$  far below the acceptable range, and the  $S_0$  value was rejected as (4) incorrect. Furthermore, the difference in concentration dependence between sonicates made with ~~an~~ <sup>and</sup> without AET does not suggest that there is a difference in the relation between  $S_0$  and  $M$  for the two kinds

of sonicates. The concentration dependence of  $S$  reflects the interaction properties of DNA molecules, while the infinite dilution extrapolation approximates the sedimentation constants of noninteracting DNA molecules. Thus there is no indication that the use of the empirical relationship between  $M_w$  and  $S_0$  is not valid for sonicates made without AET.

## APPENDIX 5

## Symmetry properties of DNA bands

The degree of symmetry or skewness of a DNA band may be evaluated by comparing the two halves of the band formed by a line dropped from the band mode  $M$  to the baseline. We designate the half band closest to the rotor center by (-), and the one farthest from the center by (+). If  $\rho_M$  is the CsCl density at  $M$ , the second moments around the band mode are defined as follows.

$$m_2(-) = \frac{\int_{-\infty}^{\rho_M} (\rho - \rho_M)^2 C(\rho) d\rho}{\int_{-\infty}^{\rho_M} C(\rho) d\rho} g^2 \text{cm}^{-6},$$

$$m_2(+) = \frac{\int_{\rho_M}^{\infty} (\rho - \rho_M)^2 C(\rho) d\rho}{\int_{\rho_M}^{\infty} C(\rho) d\rho} g^2 \text{cm}^{-6}$$

Values of  $m_2(-)$  and  $m_2(+)$  have been calculated for the bands of samples I-IV and are presented below, with values of  $\sigma_T^2$  obtained from the

distributions. The relative areas

$$R(-) = \frac{A(-)}{A(-) + A(+)} ,$$

$$R(+) = \frac{A(+)}{A(-) + A(+)}$$

are also tabulated.

Sample	R(-)	R(+)	$m_2(-) \times 10^6$	$m_2(+) \times 10^6$	$\sigma_T^2 \times 10^6$
I	0.53	0.47	10.1	4.0	7
II	0.50	0.50	11	12.5	11.8
III	0.45	0.55	37	56	47
IV	0.50	0.50	57	58	57

A distinct skewness in the direction of decreased density has been observed in bands of DNA prepared from E. coli K12 by the method of Marmur. No similar asymmetry has been observed in bands of DNA isolated from E. coli B by dodecyl sulfate lysis and density gradient centrifugation. Because this skewness is so rapidly reduced by sonication of the DNA, it is unlikely that it is due entirely to differences in nucleotide composition among the molecules. If only 5% of the DNA molecules contained of the order of 0.005 weight fraction of protein, a skewness similar to that observed could be produced.

With the exception of sample III, samples of sonicated DNA formed symmetrical bands with  $m_2(-) \approx m_2(+)$   $\approx \bar{\sigma}_T^2$ . Examination of the band shape for sample III (fig. 9) reveals that the skewness consists of a "tail" of the (+) half band, resulting in an increase in  $m_2(+)$ . The "tail" appears to be due to a small broad band centered at a position in the cell where the CsCl density is approximately  $0.015 \text{ g cm}^{-3}$  greater than at the band mode. A skewness similar to that observed in this case could be produced if less than 10% of the DNA sample had been denatured.

## TABLES

Table 1 Physical parameters of E. coli K12 DNA, samples I-IV

Data from sedimentation velocity and density gradient analysis of E. coli K12 DNA are tabulated. The following parameters are included:

$S_0$	the infinite dilution value of the sedimentation coefficient $S_{20,w}^{50}$ in Svedberg units.
$M_w, M_n$	the weight average and number average molecular weight; $Cs^+$ refers to the cesium salt of DNA, $Na^+$ to the sodium salt.
$\sigma_T^{-2}$	the variance of the density gradient distribution in $g^2 cm^{-6}$ .
$\sigma_{M_n}^{-2}$	the contribution of thermal motion to $\sigma_T^{-2}$ in $g^2 cm^{-6}$ .
$\sigma_{GC}^{-2}$	the contribution of nucleotide compositional heterogeneity to $\sigma_T^{-2}$ in $g^2 cm^{-6}$ .
k	the average number of scissions per molecule of the sonicate, with respect to the original sample.

Table 1:

Sample	I	II	III	IV
Sonication (minutes)	0	1	2	4
AET	0	0	0	4
$S_o$	31	21.5	11.6	9.8
$M_w^{Na^+} \times 10^{-6}$	15	6.5	1.5	1.0
$M_w/M_n$	1.1	2.1	1.6	1.3
$M_n^{Cs^+} \times 10^{-6}$	18.5	4.2	1.3	1.0
k	0	3.4	13	17.5
$\bar{\sigma}_{M_n}^{-2} \times 10^6$	1.9	8.4	27	35
$\bar{\sigma}_T^{-2} \times 10^6$	4.0 <sup>A</sup>	11.8	37 <sup>B</sup>	57
$\bar{\sigma}_{GC}^{-2} \times 10^6$	2.1	3.4	10	22
$\bar{\sigma}_{GC}/\bar{\sigma}_{M_n}$	1.0	.62	.62	.80

A  $m_2(+)$  is tabulated, as  $\bar{\sigma}_T^{-2}$  probably includes the effect of traces of protein on  $m_2(-)$ , as described in Appendix 5.

B  $m_2(-)$  is tabulated, as  $\bar{\sigma}_T^{-2}$  probably includes the effect of denatured DNA, as described in Appendix 5.

**Table 2** Physical parameters of E. coli K12 transfer DNA,  
samples V-VII

Data from sedimentation velocity and density gradient analysis of E. coli K12 transfer DNA are tabulated. Parameters are as defined for table 1, except that the variances have been estimated from second moments of half bands around band modes as follows

$$\sigma_T^{-2} = m_2(U(-)) \quad \text{for fragments of unlabelled DNA}$$

$$\sigma_T^{-2} = m_2(H(+)) \quad \text{for fragments of hybrid DNA}$$

The validity of this procedure has been discussed in the text and in Appendix 5.

Table 2:

Sample	V	VI	VII
Sonication (minutes)	0	1	4
AET	0	0	0
$S_o$	27.8		11.9
$M_w \times 10^{-6}$	12		1.6
$M_w/M_n$	1.2		1.3
$M_n^{Cs^+} \times 10^{-6}$	13.6		1.7
k	0	$6^A$	7
$\sigma_{M_n}^{-2} \times 10^6$	2.6		21
$\sigma_T^{-2} \times 10^6$ U(-)	9.6	29	33
H(+)	6.4	28	31
Separation of modes $g\text{ cm}^{-3}$	0.0165	0.0161	0.0154
Ratio of areas U(-)/H(+)	1.44	1.34	1.32
Ratio of heights U(-)/H(+)	1.23	1.22	1.26

A k estimated from the variance of the unlabelled DNA band.

**FIGURES**

**Figure 1 -** Density gradient equilibrium distributions for fragments of hybrid DNA, calculated from two models for the molecular arrangement of the subunits.

Equilibrium density gradient distributions have been calculated from the side-to-side model for the arrangement of the subunits and from the end-to-end model, for the following values of  $k$ , the average number of scissions per molecule:  $k = 0, 0.5, 2, 8$ . DNA concentrations relative to the peak concentration of the initial band  $\frac{C(k)}{C_0(0)}$  are plotted against density distance from the band mean, relative to the square root of initial band variance  $\rho - \rho_0 / \bar{\sigma}_T(0)$ .

Details of the calculation are given elsewhere (Appendix 1).

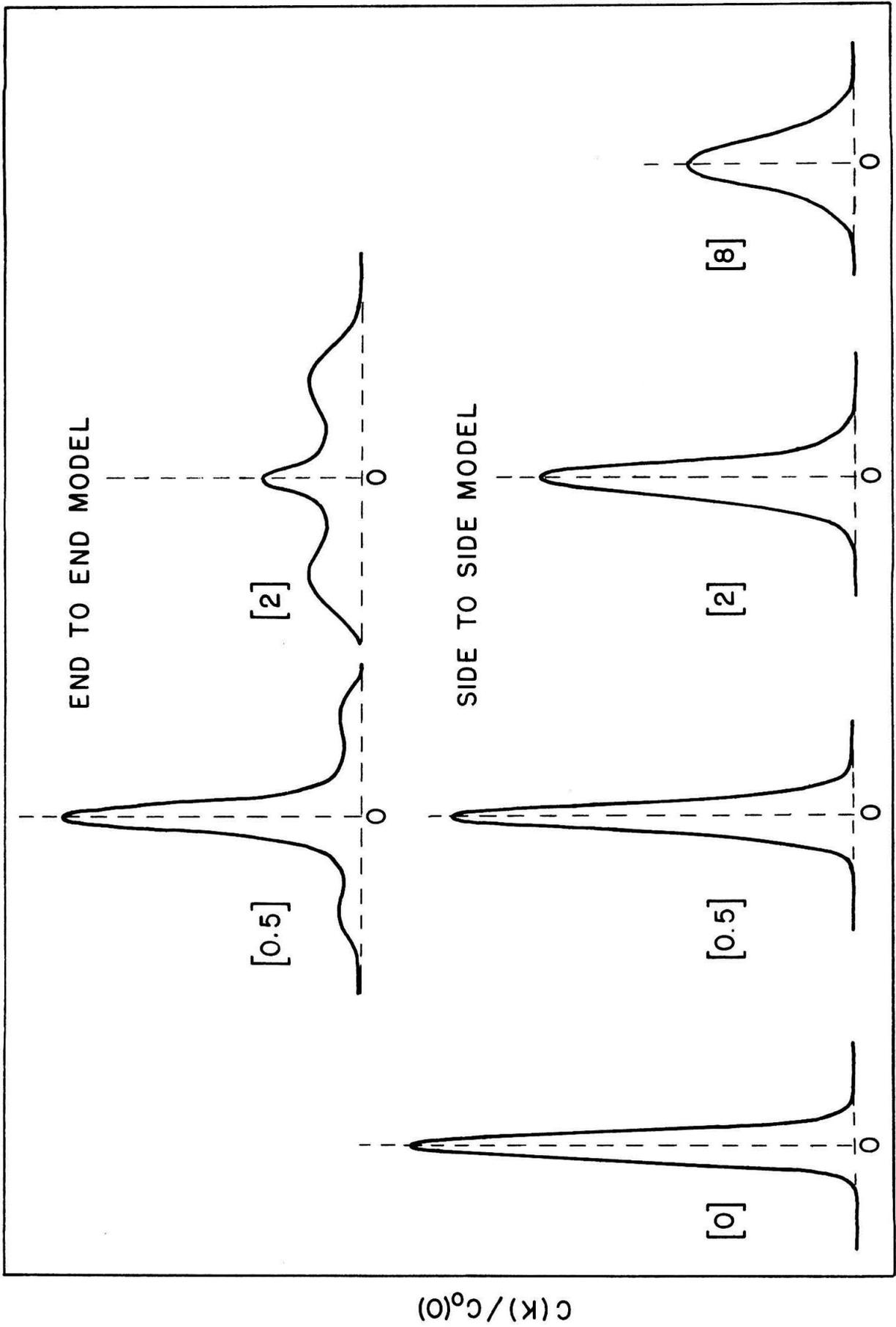
The following assumptions were made.

- 1) Fragmentation is random, occurring at discrete breaking points along the length of the molecules.
- 2) The density separation between unlabelled and fully labelled fragments is given by

$$\rho_L - \rho_U = 16 \bar{\sigma}_T(0)$$

This approximates the density separation found by experiment.

- 3) The molecules are homogeneous in nucleotide composition. This assumption is essentially equivalent to the assumption that base pairs are randomly distributed (Appendix 1).



$$(p-p_0) / \bar{\sigma}_{M_n}(0)$$

Fig. 1

Figures 2 and 3 - Comparison of experimental and calculated equilibrium distributions for fragments of hybrid DNA.

Hybrid  $^{13}\text{C}$ ,  $^{15}\text{N}$  DNA was isolated from sodium dodecyl sulfate lysates of E. coli B by preparative density gradient centrifugation. Fragmentation was accomplished by vigorous manual shaking of DNA-CsCl solutions. Equilibrium distributions for the initial sample (D195) and of those successively derived from it by shaking (D198, D200, D202) are compared with distributions calculated from the two models for the arrangement of the subunits. All distributions are plotted as the logarithm of the DNA concentration relative to the peak concentration  $\log(C/C_0)$ , versus the squared density distance from the band mean, relative to the variance of the initial distribution  $(\rho - \rho_0)^2 / \bar{\sigma}_T^2(0)$ .

In figure 2 the experimental distributions are superimposed upon a family of distributions calculated from the side-to-side model for the arrangements of the subunits. The value of  $k$ , the average number of scissions per molecule, is given for each calculated distribution.

In figure 3, the experimental distributions are superimposed upon a family of distributions calculated from the end-to-end model. Details of the calculations are given elsewhere (Appendix 1).

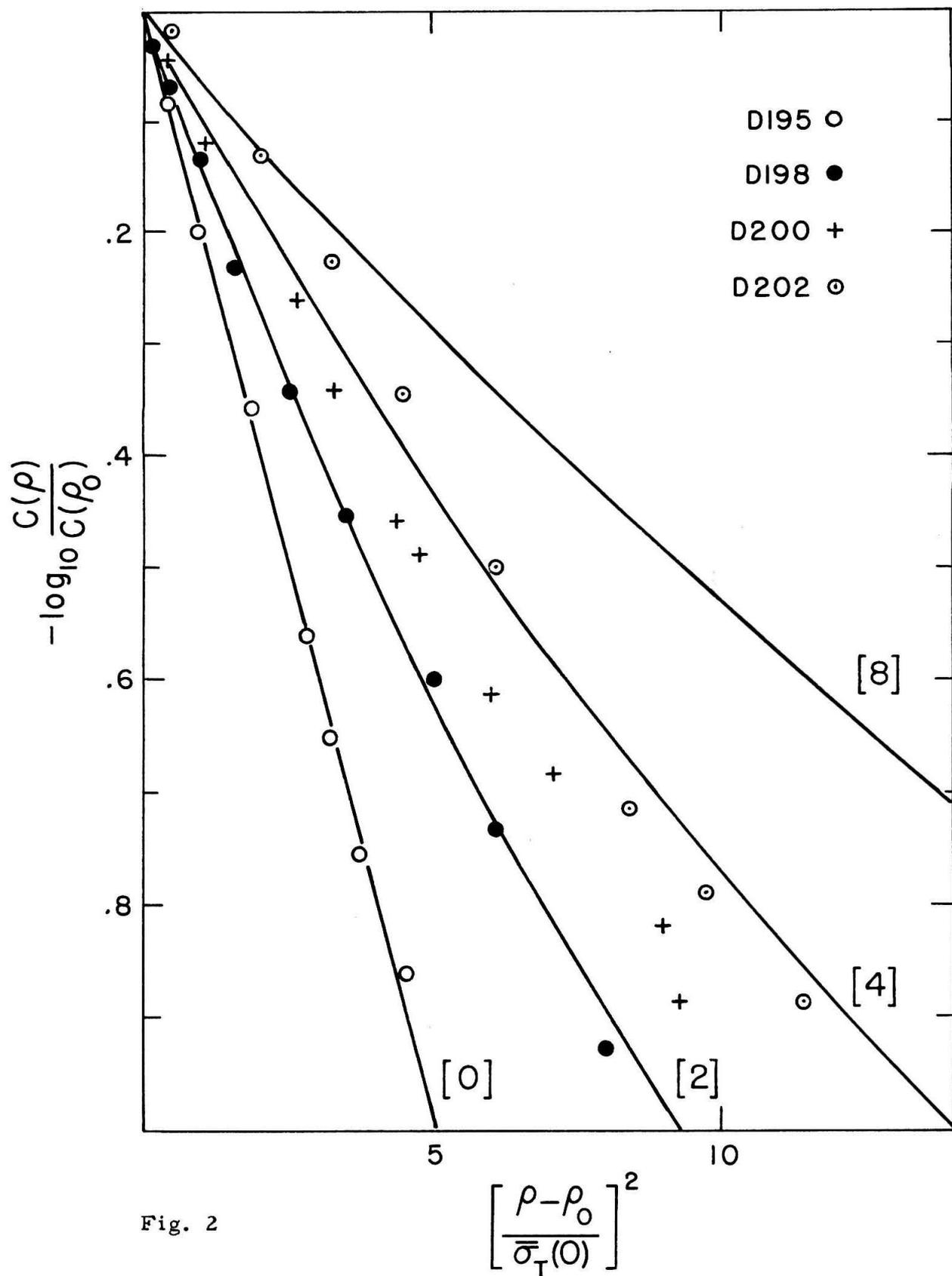


Fig. 2

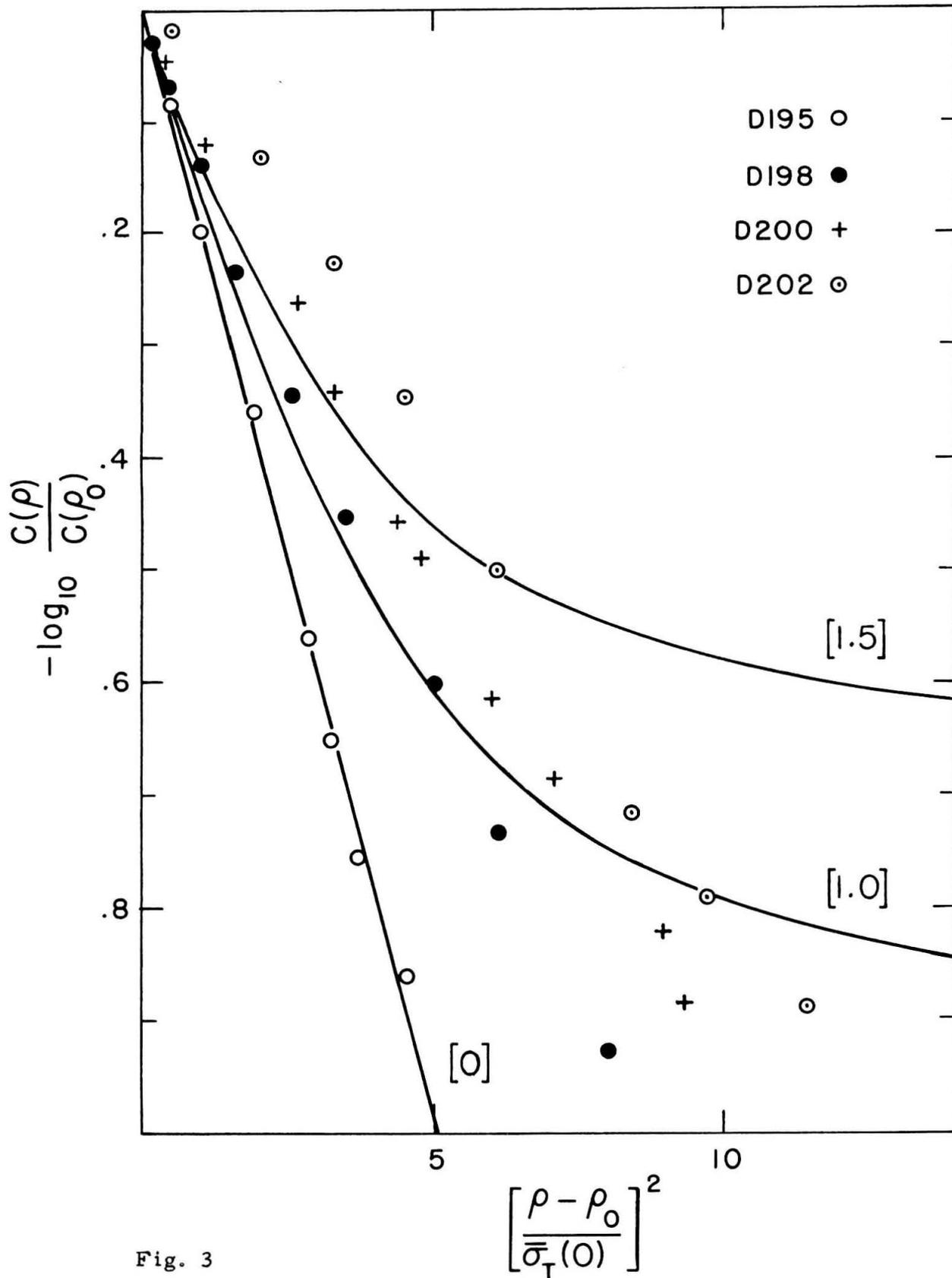


Fig. 3

Figure 4 - Electron micrographs of DNA from E. coli K12

Solutions of DNA in ammonium acetate - ammonium carbonate buffer were sprayed onto the surface of freshly cleaved mica with a low pressure atomizer. The DNA was shadow cast with platinum at a shadow to height ratio of 5: 1. After backing the platinum with a carbon substrate, it was floated off the mica and mounted on grids for electron microscopy. Polystyrene latex spheres 0.25 microns in diameter were included in the sample to aid in focussing. Magnification 20,000 X.

- a) Unsonicated E. coli K12 DNA (Sample I). Branching of molecules indicates side-by-side aggregation, and has been attributed to traces of nonvolatile salts in the preparation (9).
- b) Sonic fragments of E. coli K12 DNA (Sample IV). The platinum layer is too thick to permit optimal visualization.

65

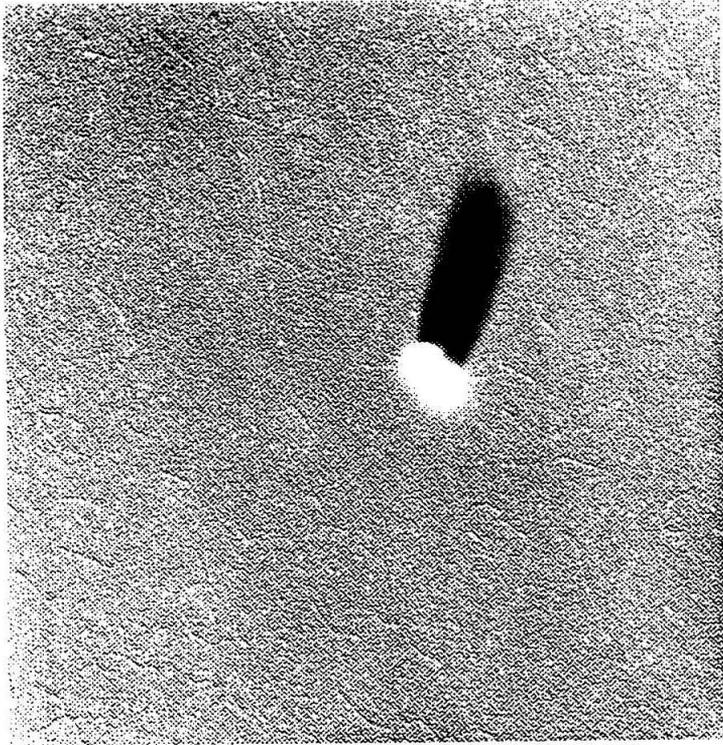


Fig 4b

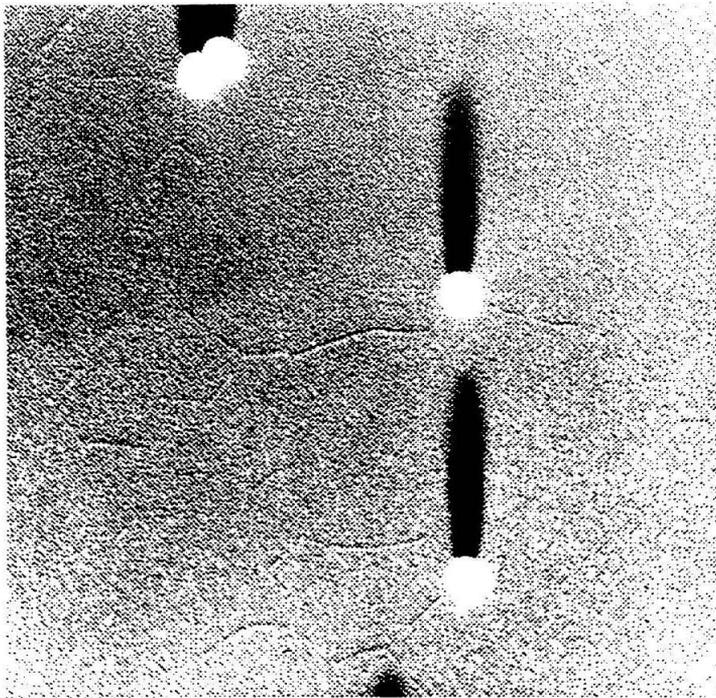


Fig 4a

**Figure 5 -** The relationship between weight average molecular weight and the infinite dilution value of the sedimentation coefficient for DNA, according to Eigner (4).  
a)

Weight average molecular weights  $M_w$  for samples of DNA have been determined by light scattering, indicated by circles, or from viscosity and sedimentation measurements, indicated by triangles. They are plotted against the infinite dilution value of the sedimentation coefficient  $S_0$ . The calibration is discussed in Appendix 3, and the following classes of samples are included.

Designation	Source of DNA	Treatment	Investigator
○	calf thymus	sonicated without AET	Doty, McGill and Rice(16)
●	D. pneumoniae	sonicated with AET	Litt (12)
▲	"	" " "	Eigner (4)
△	E. coli	unfragmented	Eigner (4)
○	calf thymus	"	Doty, McGill and Rice(16)
○	salmon sperm	"	Geiduschek and Holtzer(26)
○	D. pneumoniae	shear fragmented and	Cavalieri and Rosenberg(17)
	"	unfragmented	Litt(12)

**Figure 5 -** The relationship shown in figure 5a, plotted without the data points. A possible alternative relationship for sonicates made without AET is shown as a dashed line, and discussed in Appendix 4.  
b)

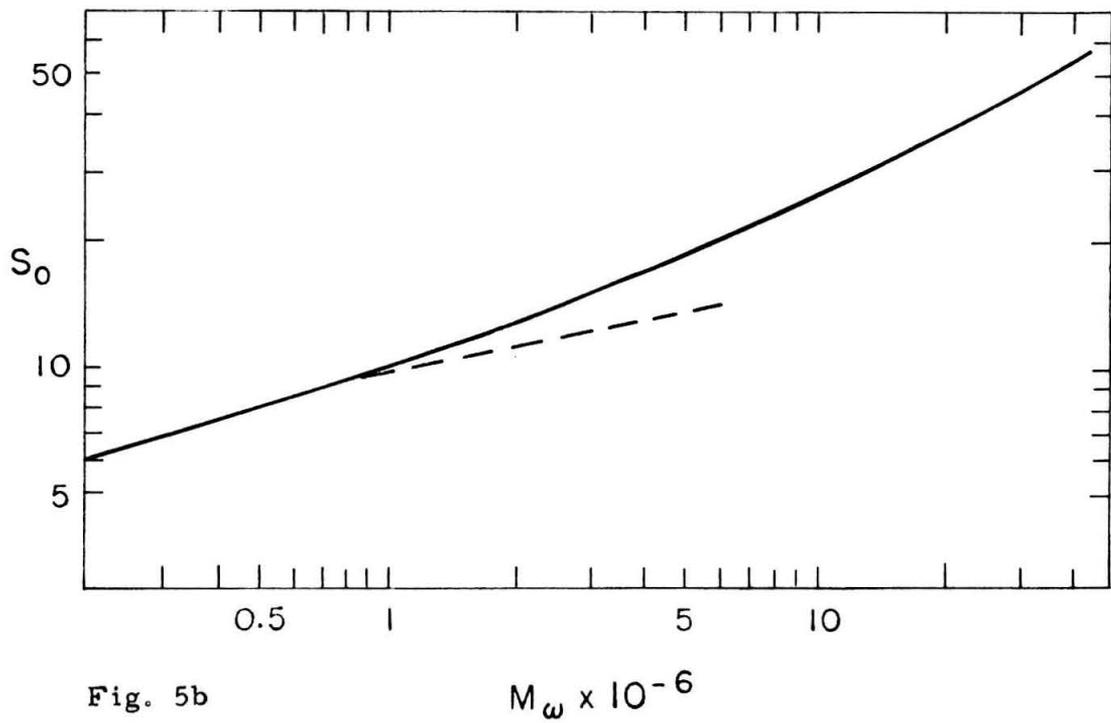
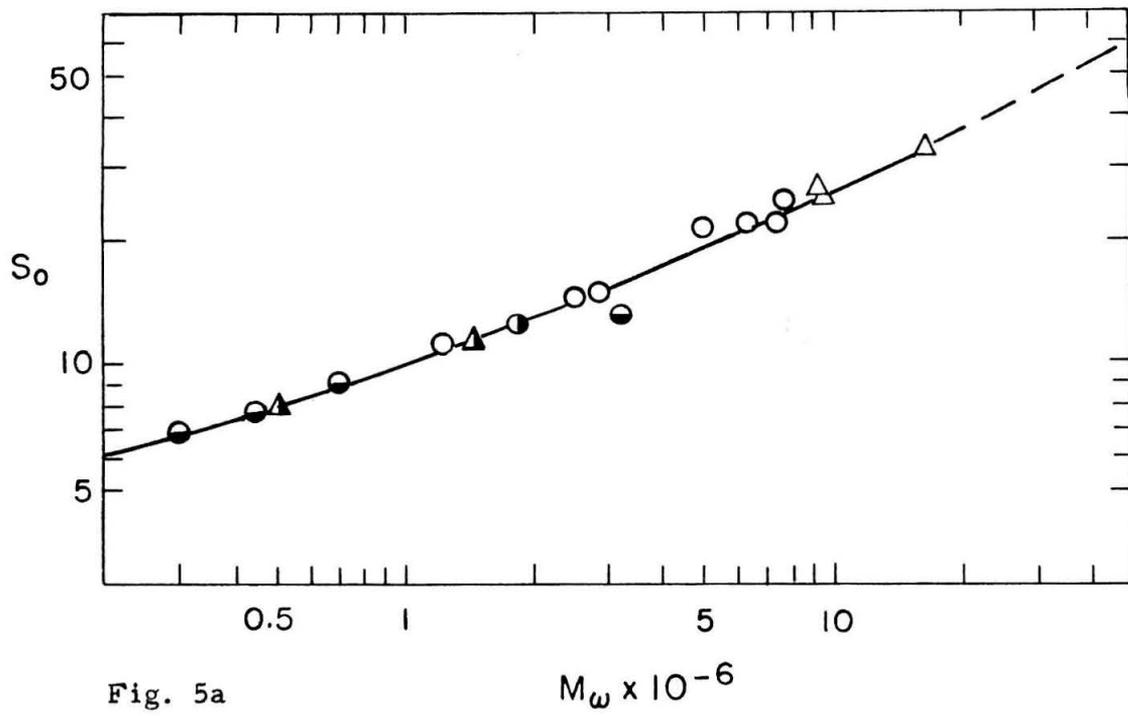


Figure 6 - The concentration dependence of the sedimentation coefficient for E. coli K12 DNA, samples I-IV.

The sedimentation coefficient ( $S_{20,w}^{50}$ ) is plotted against DNA concentration ( $\mu\text{g/ml}$ ) for four samples of E. coli K12 DNA.

- |     |                                      |
|-----|--------------------------------------|
| I   | original sample, unsonicated         |
| II  | " " , sonicated 1 minute without AET |
| III | " " , " 2 minutes without AET        |
| IV  | " " , " 4 minutes with AET           |

No consistent differences in  $S$  were noted when runs at rotor speeds of 23,150 RPM and 44,770 RPM were compared.

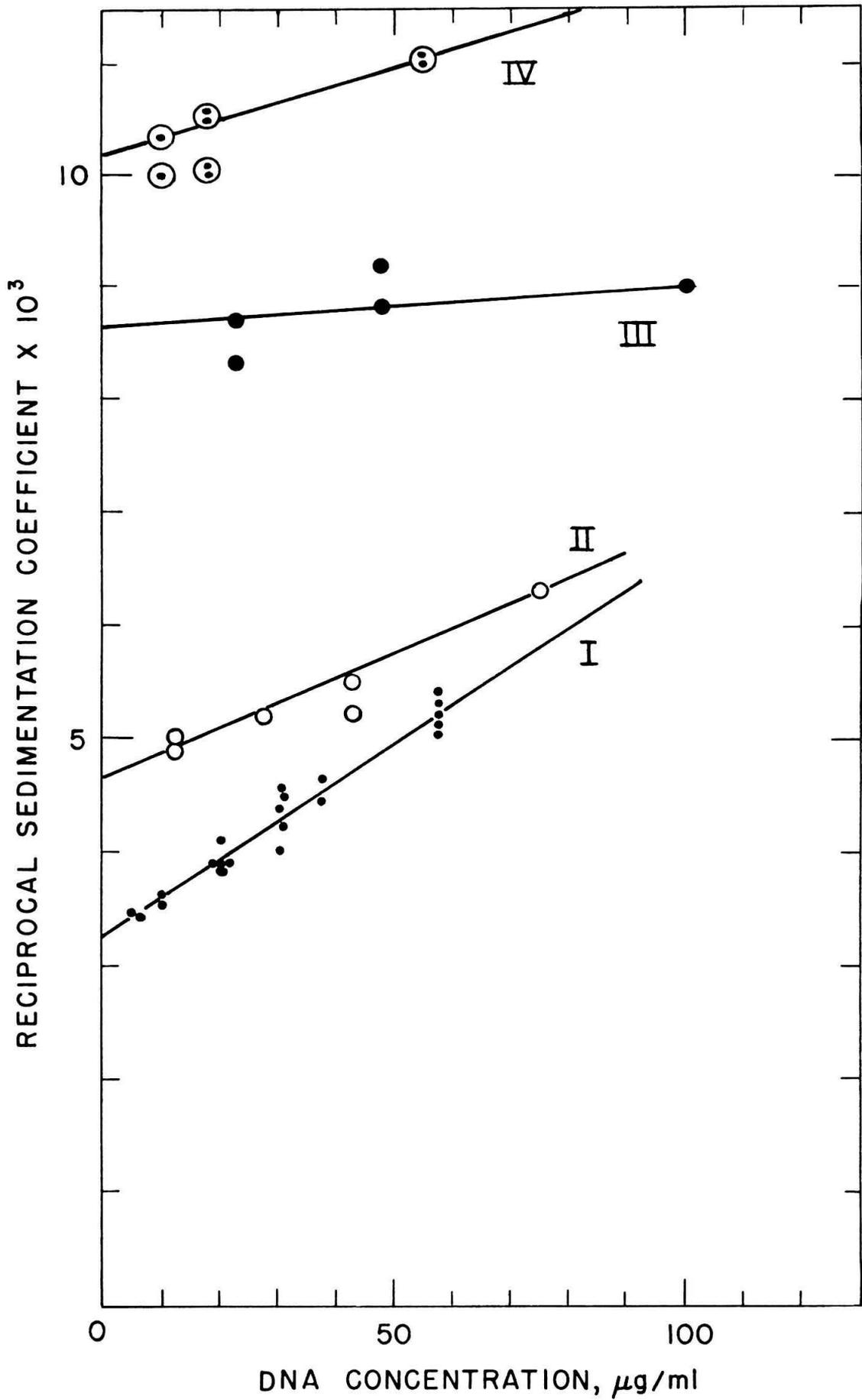


Fig. 6

Figure 7 - Integral distribution of sedimentation coefficients for E. coli K12 DNA, samples I-IV.

The weight fraction of DNA with a sedimentation coefficient  $S_0^1$  less than the sedimentation coefficient  $S_0$  is plotted against  $S_0$ . Distributions were obtained from sedimenting boundary profiles, and were essentially independent of centrifugation time. Pertinent data are as follows:

Sample	I	II	III	IV
Run	B1094-1	C849-1	C857-1	D442-2
DNA concentration ( $\mu\text{g/ml}$ )	12	12	22	10
Rotor speed (RPM)	35,600	23,150	23,150	35,600
Effective time of centrifugation (minutes)	23	49	79	41
$S_{20,w}^{50}$ (Svedberg units)	27	20	11.5	9.5
Distance of boundary midpoint from meniscus(cm)	0.336	0.226	0.210	0.324

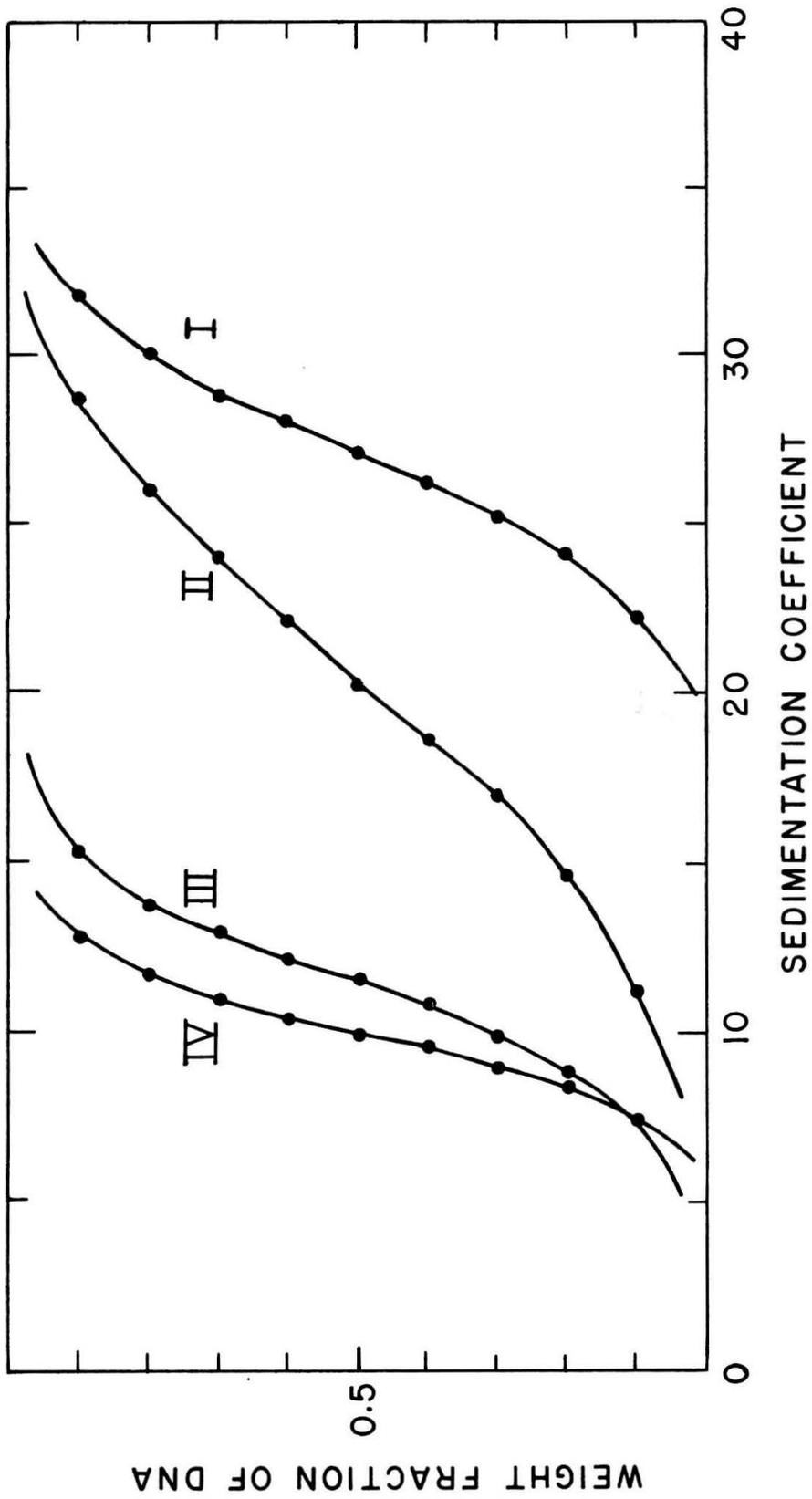


Fig. 7

Figures 8-11 - Equilibrium density-gradient distributions of E. coli K12 DNA, samples I-IV.

Distributions of DNA concentration with respect to distance from rotor center for samples I-IV after centrifugation to equilibrium in 7.7 M CsCl. Pertinent data regarding each distribution are as follows:

Sample	I	II	III	IV
Figure	8	9	10	11
Photograph	D414-3	D425-4	D451-4	D417-5
Rotor speed (RPM)	35,600	35,600	44,770	44,770
Density of CsCl solution ( $\text{g cm}^{-3}$ )	1.73	1.71	1.71	1.71
Amount of DNA ( $\mu\text{g}$ )	2	2.4	1.3	4
Time of centrifugation (hours)	60	72	63	60
t* + 8 (hours) <	56	68	50	56

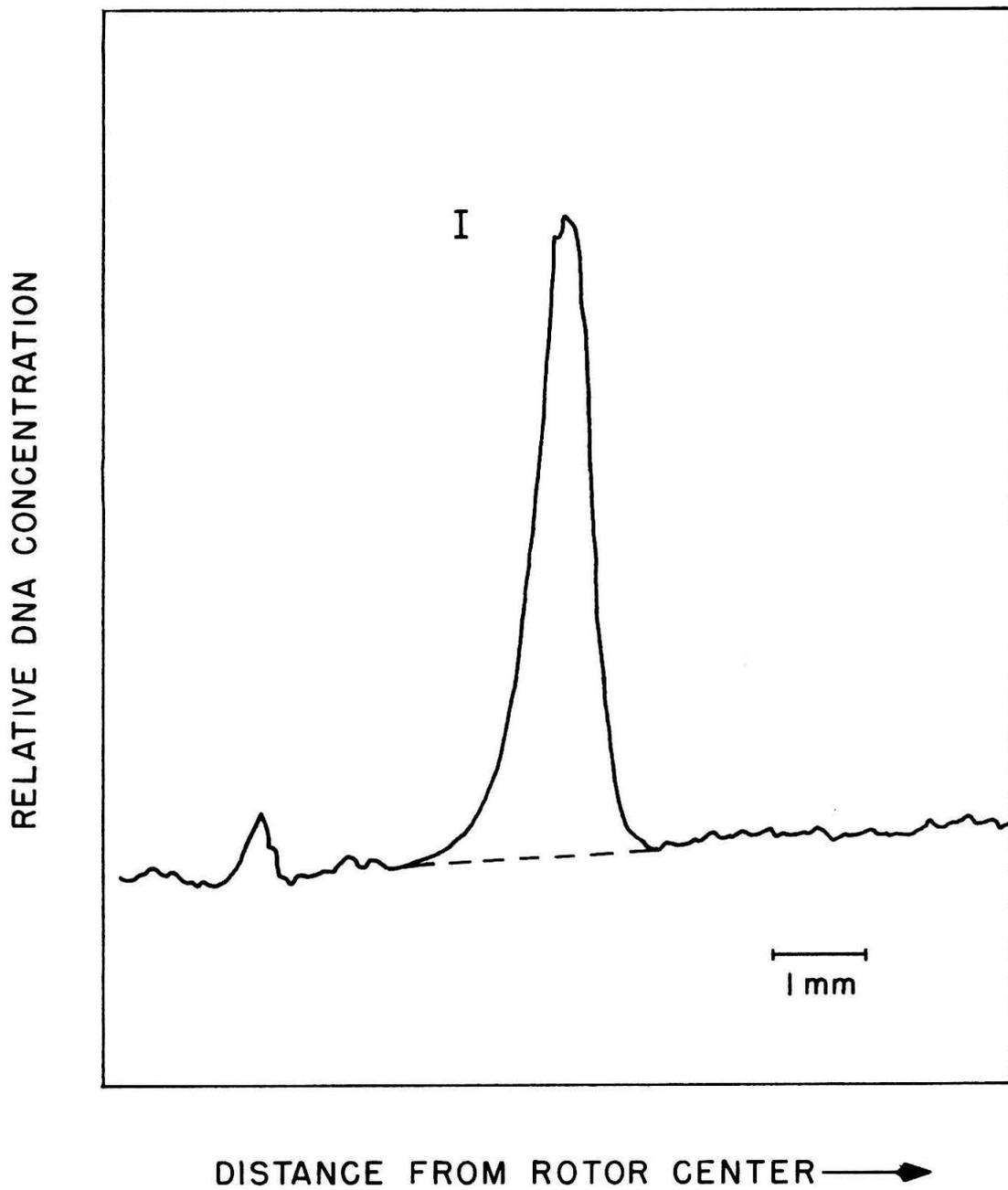


Fig. 8

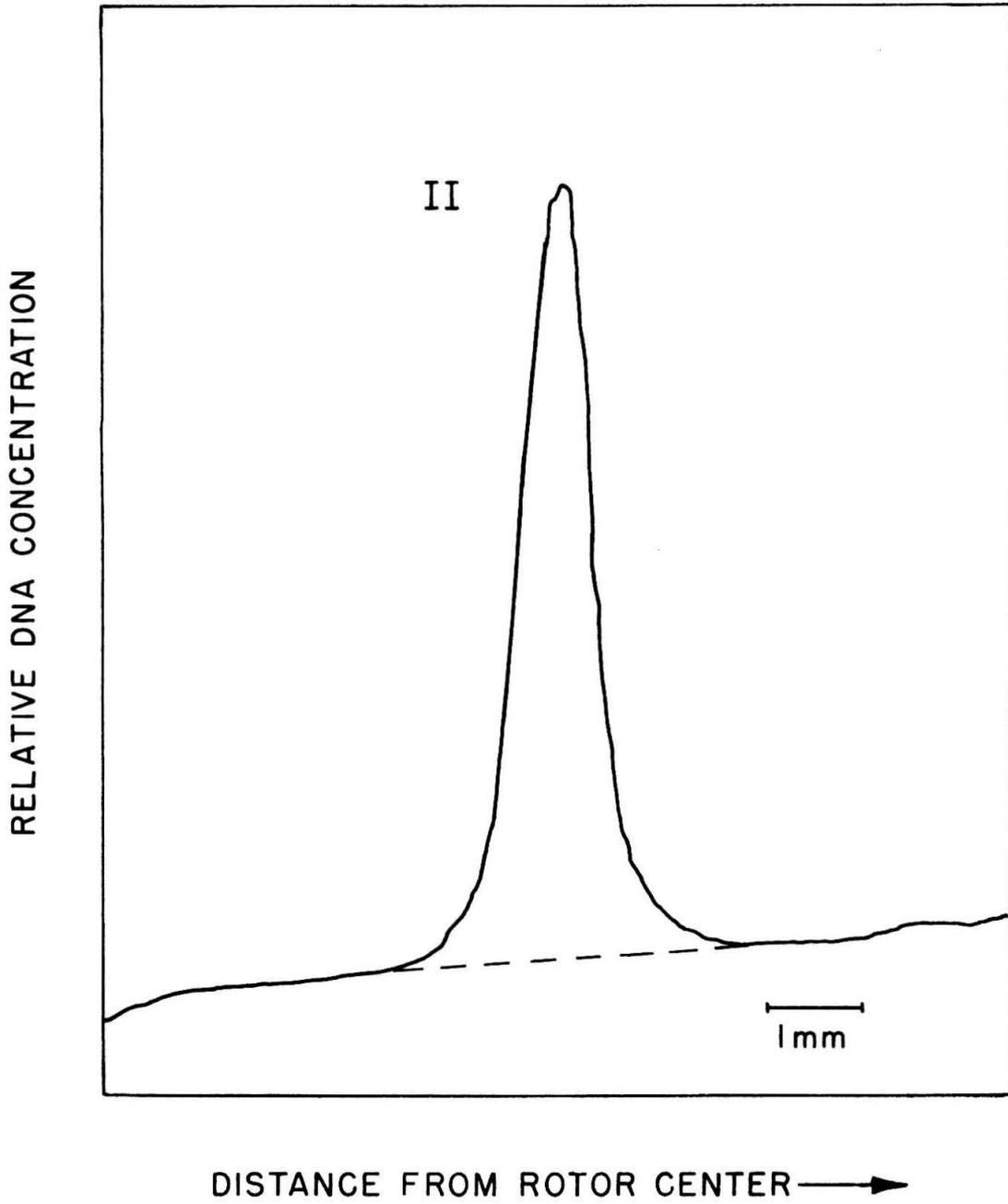


Fig. 9

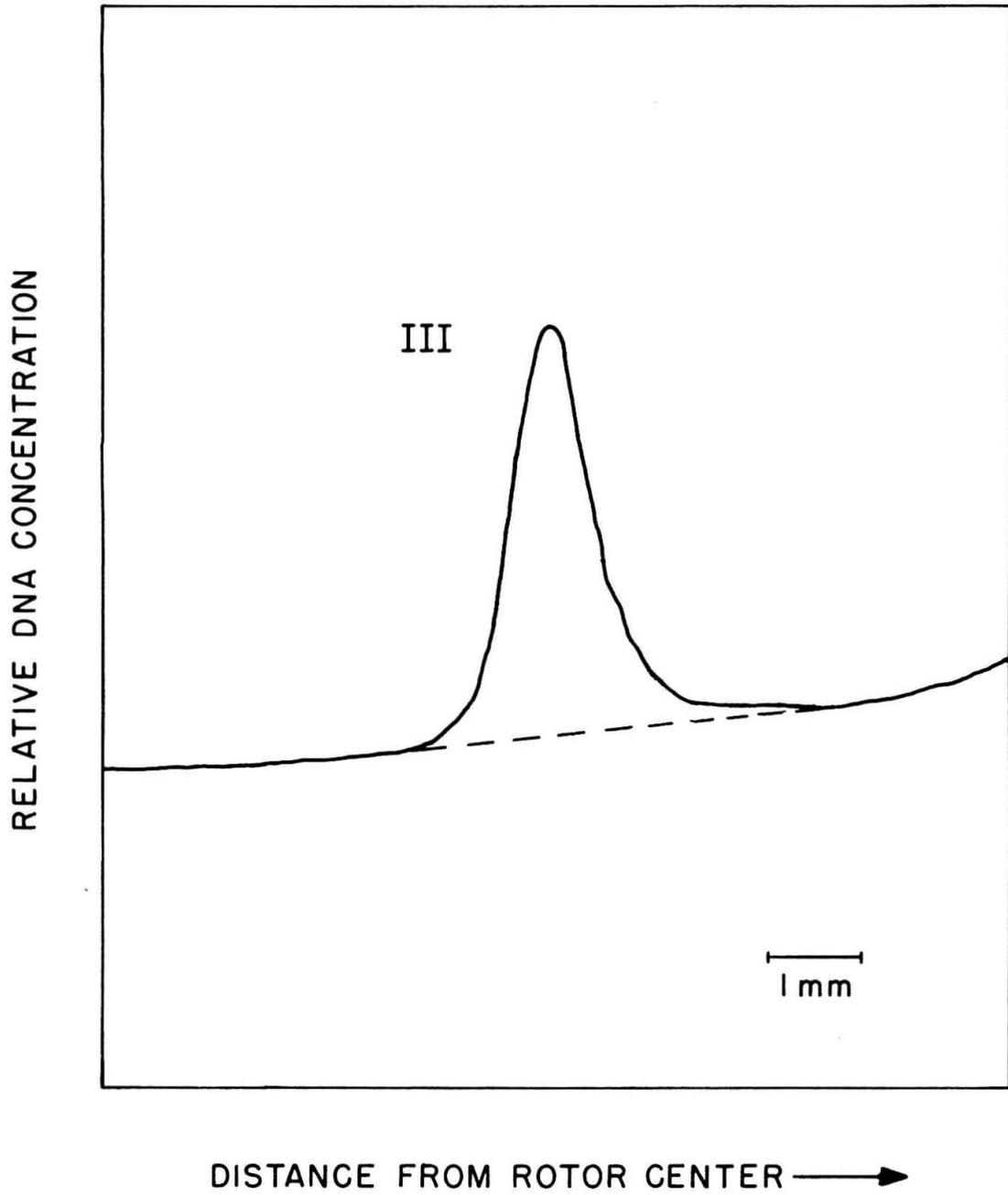


Fig. 10

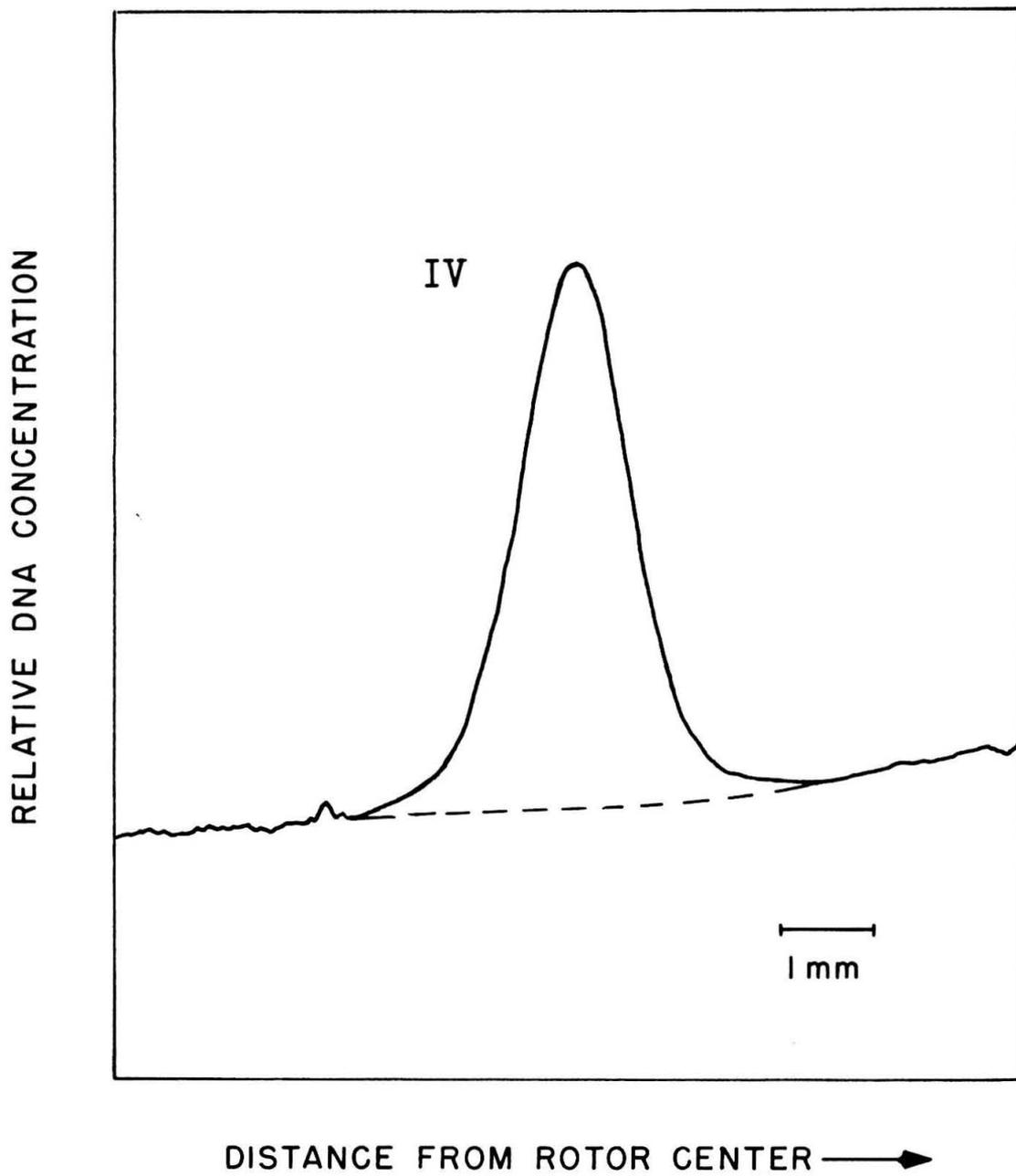


Fig. 11

Figure 12 - The concentration dependence of the sedimentation coefficients for E. coli K12 transfer DNA, samples V, VII

The sedimentation coefficient  $S_{20,w}^{50}$  is plotted against DNA concentration for two samples of E. coli K12 transfer DNA.

- V original transfer DNA, unsonicated  
 VII " " " , sonicated 4' without AET

The dotted lines refer to samples I-IV, figure 3.

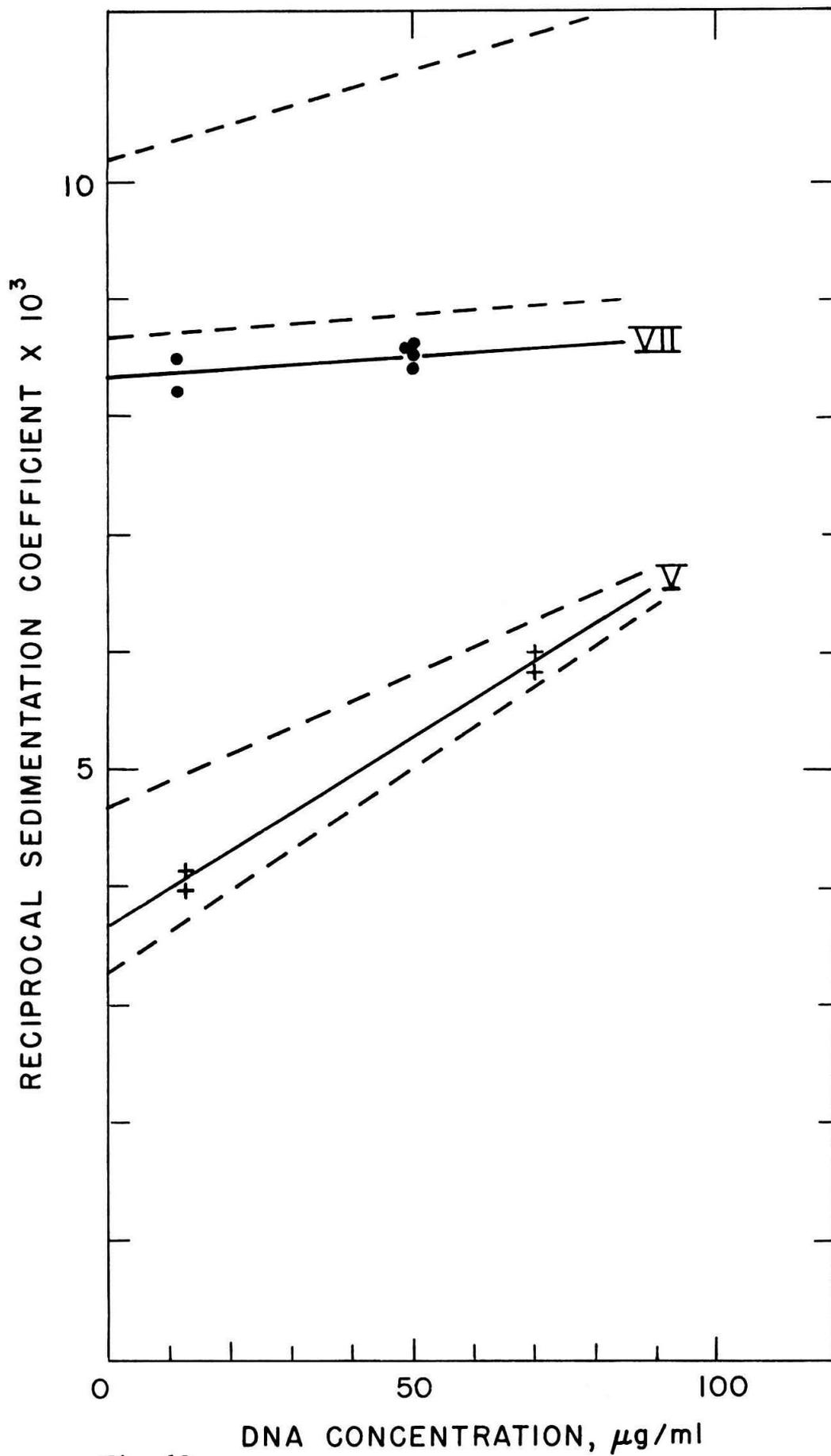


Fig. 12

Figure 13 - Integral distributions of sedimentation coefficients for E. coli K12 transfer DNA, samples V, VII.

The weight fraction of DNA with a sedimentation coefficient  $S_0^1$  less than the sedimentation coefficient  $S_0$  is plotted against  $S_0$ . Distributions were obtained from sedimenting boundary profiles and were essentially independent of centrifugation time. Pertinent data are as follows:

Sample	V	VII
Run	A2040-1	C906
DNA concentration ( $\mu\text{g/ml}$ )	12	12
Rotor speed (RPM)	23,150	44,770
Effective time of centrifugation (minutes)	61	34
$S_{20,w}^{50}$ (Svedberg units)	24.4	11.6
Distance of boundary midpoint from meniscus(cm)	.342	.342

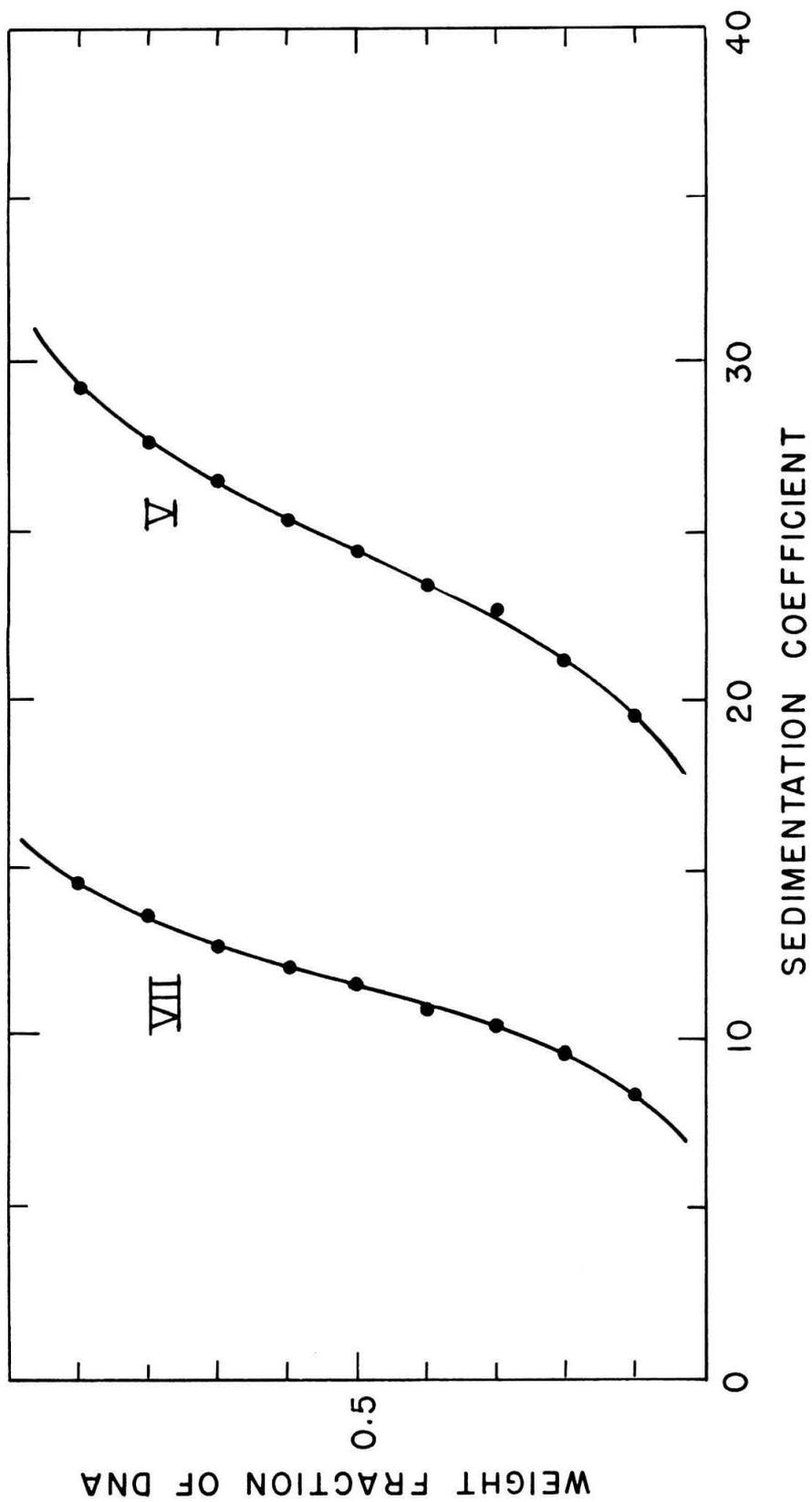


Fig. 13

Figures 14, 15 and 16 - Equilibrium density gradient distributions of *E. coli* R12 transfer DNA, samples V, VI and VII.

Distributions of DNA concentration with respect to distance from the rotor center for samples V-VII after centrifugation to equilibrium in 7.7 M CsCl. Pertinent data regarding each distribution are as follows:

Sample	V	VI	VII
Figure	14	15	16
Photograph	D499-3	D502-4	D500-3
Rotor speed (RPM)	44,770	44,770	44,770
Density of CsCl solution (g cm <sup>-3</sup> )	1.736	1.736	1.72
Amount of DNA (μg)	1	4	5
Time of centrifugation (hours)	24	53	60
t* + 8 (hours) <	26		59

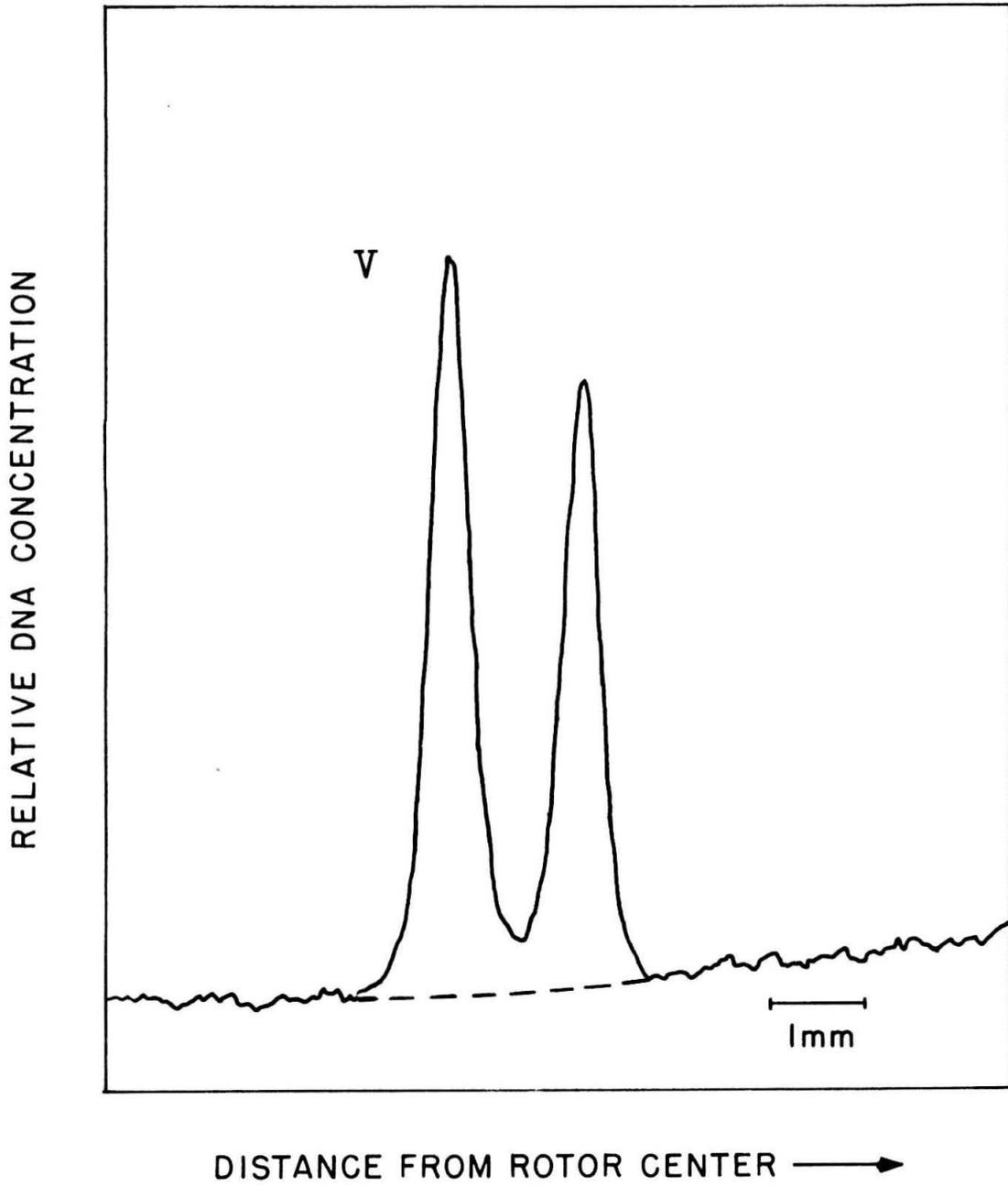


Fig. 14

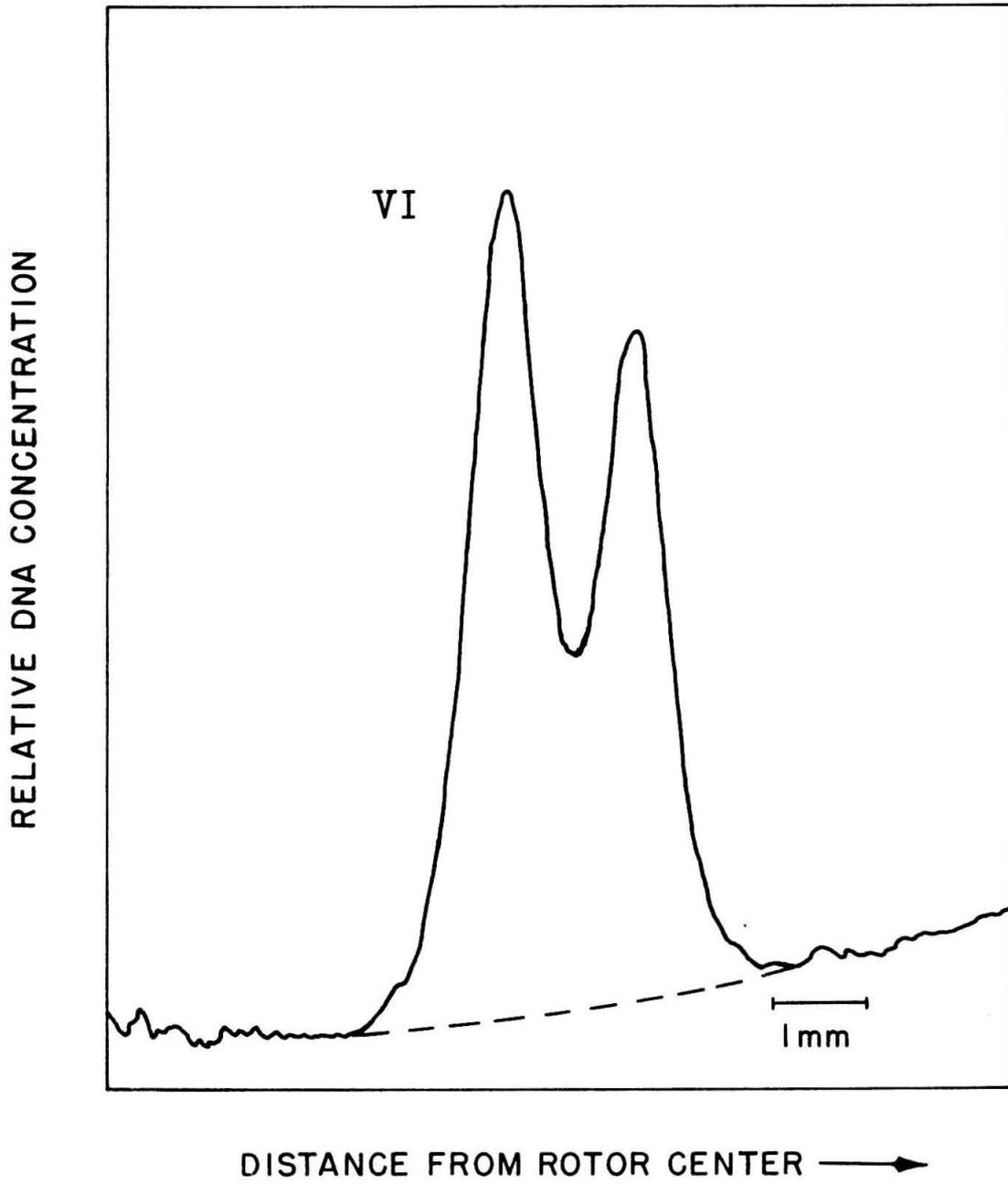


Fig. 15

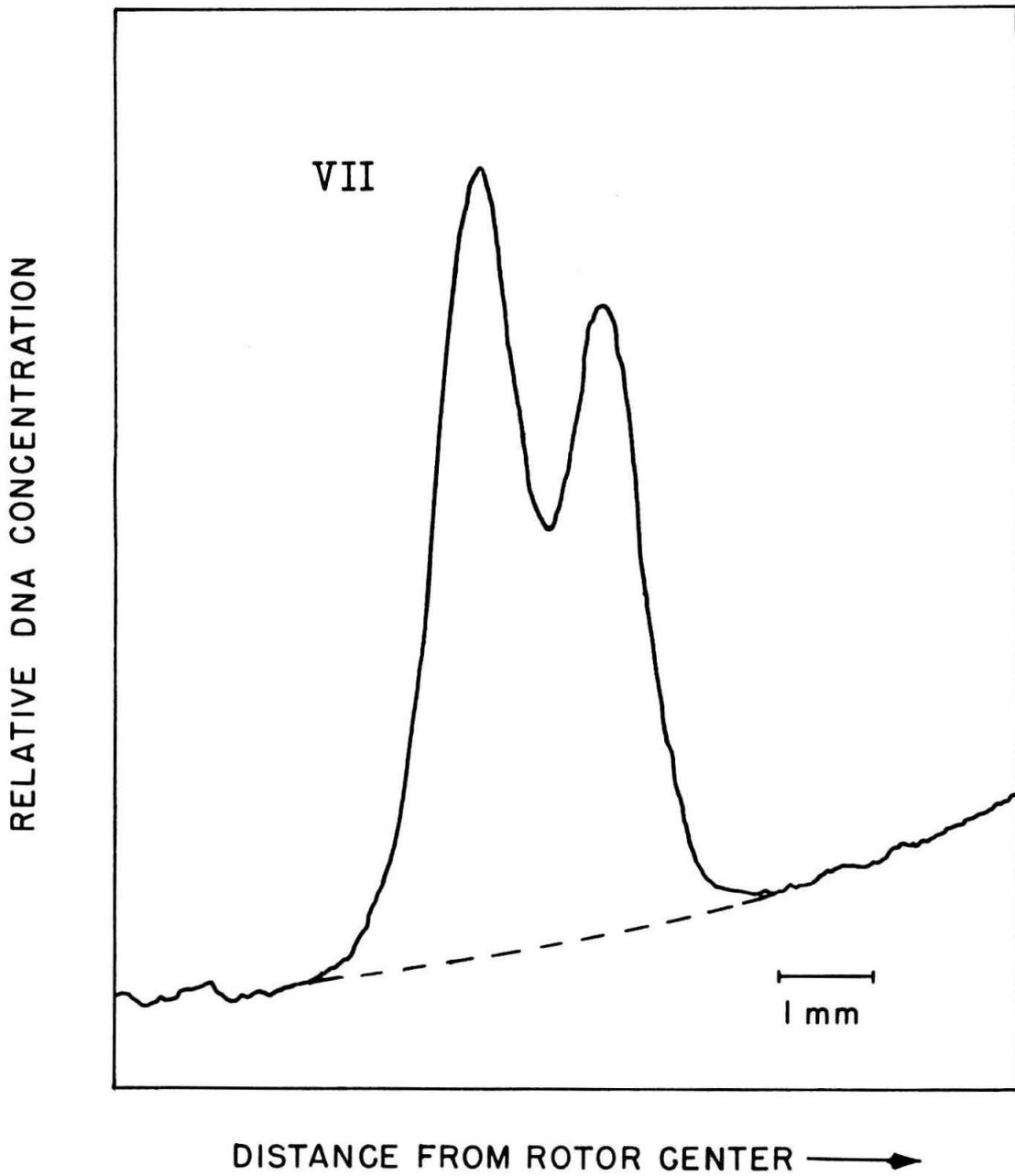


Fig. 16

Figure 17 - Contributions to the variance of equilibrium distributions of hybrid DNA fragments.

The contributions of thermal motion  $\bar{\sigma}_{M_n}^{-2}$ , nucleotide compositional heterogeneity  $\bar{\sigma}_{GC}^{-2}$  and isotopic heterogeneity  $\bar{\sigma}_I^{-2}$  to the variance of bands of hybrid DNA  $\bar{\sigma}_I^{-2}$  are compared at various stages of fragmentation. Values of  $\frac{\bar{\sigma}_{M_n}(k)}{\bar{\sigma}_{M_n}(0)}$ ,  $\frac{\bar{\sigma}_{GC}(k)}{\bar{\sigma}_{M_n}(0)}$ , and  $\frac{\bar{\sigma}_I(k)}{\bar{\sigma}_{M_n}(0)}$  have

been plotted against  $k$ , the average number of scissions per molecule.

Values of  $\frac{\bar{\sigma}_{M_n}(k)}{\bar{\sigma}_{M_n}(0)}$  were calculated from the expression

$$\frac{\bar{\sigma}_{M_n}(k)}{\bar{\sigma}_{M_n}(0)} = (k+1)^{1/2}$$

Values of  $\frac{\bar{\sigma}_{GC}(k)}{\bar{\sigma}_{M_n}(0)}$  were taken from the data of table 1. For the

calculation of  $\frac{\bar{\sigma}_I(k)}{\bar{\sigma}_{M_n}(0)}$  see text, p. 24 and Appendix 1.

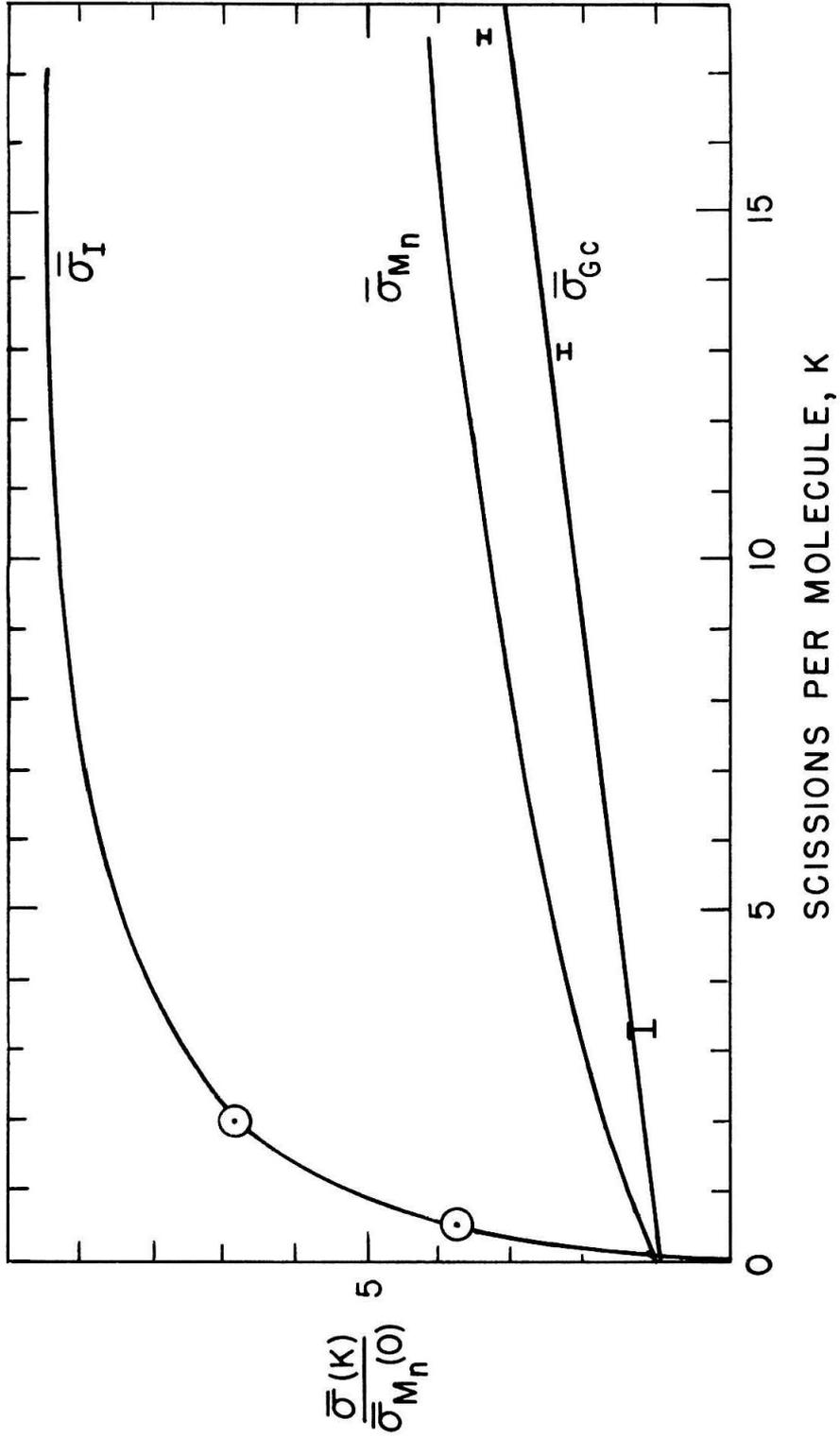


Fig. 17

Figure 13 - Comparison of observed and predicted values of  $\bar{\sigma}_T^{-2}(k)$  for fragments of hybrid DNA.

A plot of  $\bar{\sigma}_T^{-2}(k)$  against  $k$  has been made for the side-to-side model for the arrangement of the subunits and for the end-to-end model. Values of  $\bar{\sigma}_T^{-2}(k)$  were obtained from figure 17 using equation 5. Observed values of  $\bar{\sigma}_T^{-2}(k)$  for sonic fragments of hybrid DNA (samples VI and VII) are plotted as open circles. The experimental data were taken from table 2.

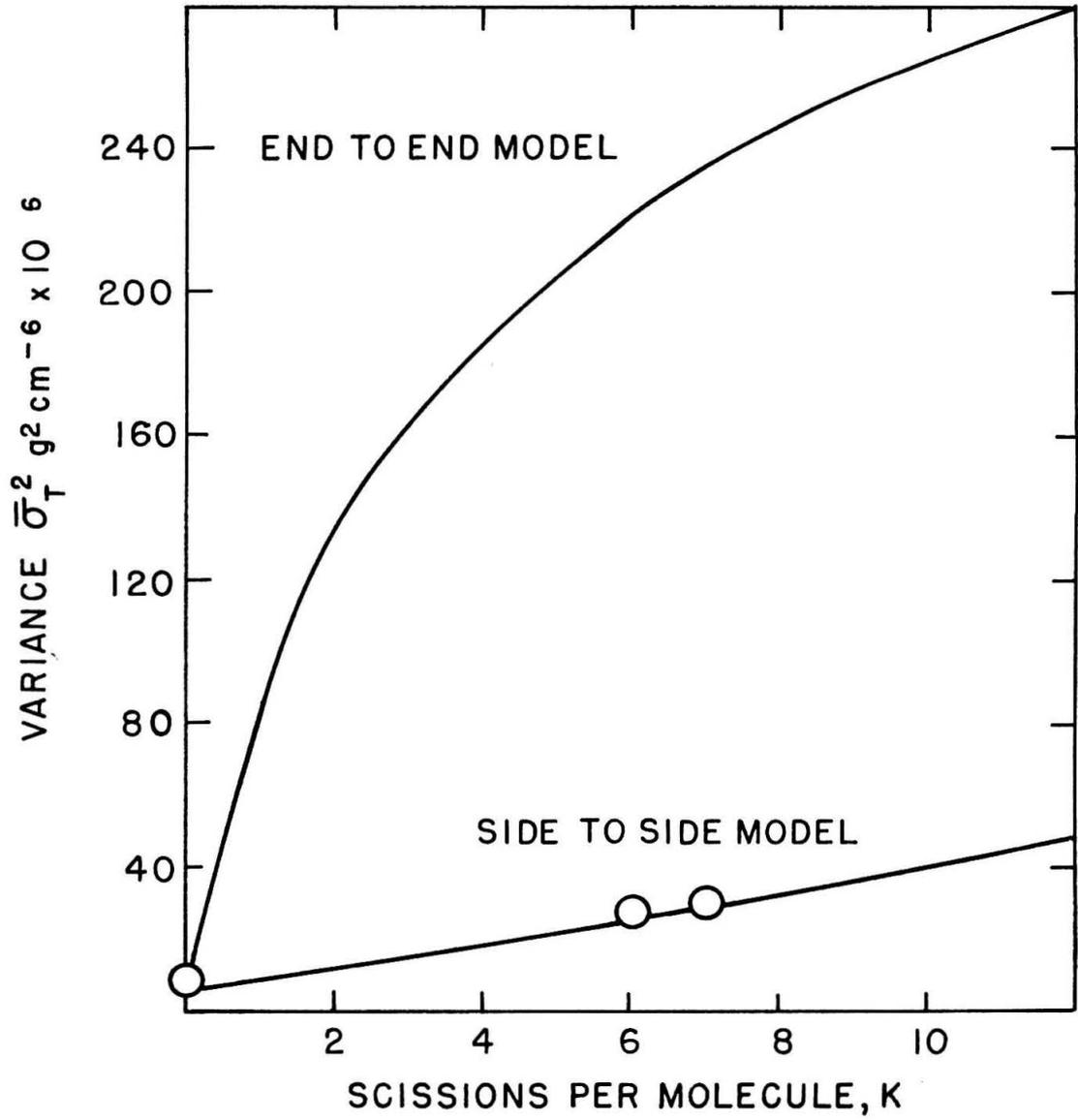


Fig. 18

References

1. M. Meselson and F. W. Stahl, PNAS, 44, 671 (1958).
2. N. Sueoka, J. Marmur and F. Doty, Nature, 183, 1429 (1959).
3. R. Rolfe and M. Meselson, PNAS, 45, 1039 (1959).
4. J. Eigner, Thesis, Harvard University (1960).
5. N. Sueoka, PNAS, 45, 1480 (1959).
6. M. Meselson, Thesis, California Institute of Technology (1957).
7. J. Marmur, submitted to J. Molecular Biology (1961).
8. C. Davern, Thesis, California Institute of Technology (1959).
9. C. E. Hall and P. Doty, JACS, 80, 1269 (1958).
10. E. Robkin, M. Meselson and J. Vinograd, JACS, 81, 1305 (1959).
11. J. Hearst and J. Vinograd, Arch. Biochem. and Biophysics, in press (1961).
12. M. Litt, Thesis, Harvard University (1958).
13. P. Davidson, PNAS, 45, 1560 (1959).
14. M. Meselson, personal communication.
15. C. E. Hall, personal communication to N. Davidson (1960).
16. P. Doty, B. B. McGill and S. A. Rice, PNAS, 44, 432 (1958).
17. L. F. Cavalieri and B. H. Rosenberg, JACS, 81, 5136 (1959).
18. L. Cavallieri, B. Rosenberg and J. Deutsch, Biochemical and Biophysical Research Communications, 1, 124 (1959).
19. R. L. Baldwin, PNAS, 45, 939 (1959).

20. H. Cramer, "Mathematical Methods of Statistics," Princeton University Press (1951), p. 180.
21. L. Mandelkern, W. R. Krigbaum, H. A. Scheraga and P. J. Flory, J. Chem. Phys., 20, #9, 1392 (1952).
22. H. A. Scheraga and L. Mandelkern, JACS, 75, 179 (1953).
23. A. Peterlin, J. Colloid Science, 10, 587 (1955).
24. J. A. V. Butler, D. J. R. Laurence, A. B. Robins and K. V. Shooter, Proc. Royal Society of London, Series A, 250, 1 (1959).
25. M. Litt, J. Marmur, H. Ephrussi-Taylor and P. Doty, PNAS, 44, 144 (1958).

## PROPOSITION I

A theory of the "renaturation" of DNA is critically examined.

The heating of DNA results in striking changes in a number of its physical, chemical, and biological properties. These changes occur abruptly when a solution of DNA reaches a characteristic temperature known as the melting temperature or the denaturation temperature. The terms "native" and "denatured" are commonly employed to differentiate the two states of DNA.

The distinguishing properties of denatured DNA of particular interest in this discussion are 1) a greatly diminished biological activity as a transforming principle (relative to the activity of native DNA), 2) an increase in the buoyant density as measured in the CsCl gradient,

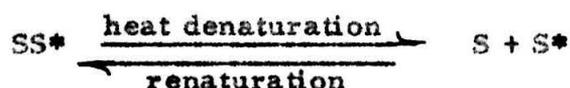
$$\rho_{\text{denatured}} = \rho_{\text{native}} + 0.016 \text{ g cm}^{-3}$$

and 3) a change in electron micrographs, in which the highly extended linear structures characteristic of native DNA are no longer observed. If DNA is first heat denatured, and then heated at a temperature below the denaturation temperature, there is a partial restoration of the biological activity, a decrease in the buoyant density from the density of denatured DNA to one near the density of native DNA.

$$\rho_{\text{renatured}} = \rho_{\text{denatured}} - 0.012 \text{ g cm}^{-3} = \rho_{\text{native}} + 0.004 \text{ g cm}^{-3}$$

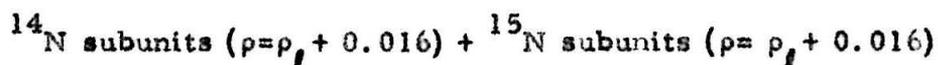
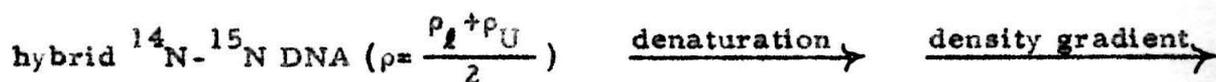
and the reappearance of the previously mentioned linear structures in electron micrographs. This restoration of properties of native DNA has been called "renaturation" (1, 2).

Doty believes that the denaturation of DNA involves the separation of the two strands of the Watson Crick helix, and that denatured DNA consists of single polynucleotide strands. In his view, renaturation is the pairing of complementary single strands to form reconstituted Watson Crick helices. If  $SS^*$  is a Watson Crick double helix,  $S$  and  $S^*$  being the complementary polynucleotide strands, Doty's proposal may be represented

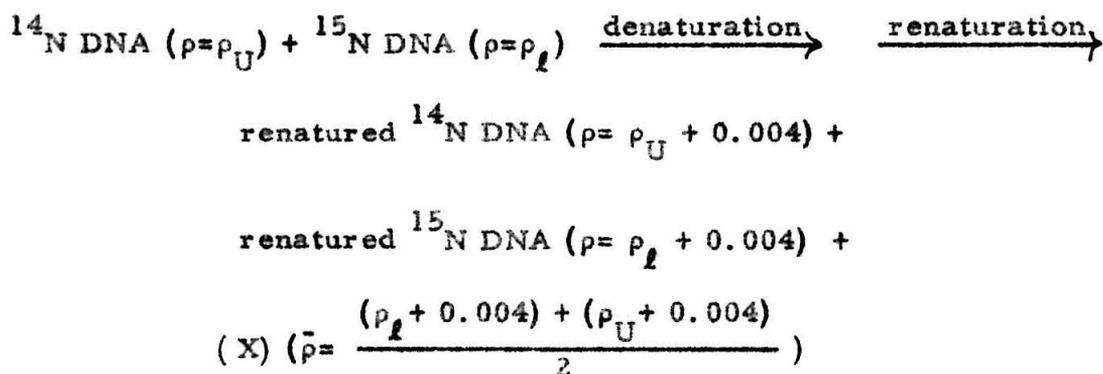


Doty bases his hypothesis on two observations described below in a) and b).

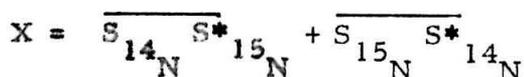
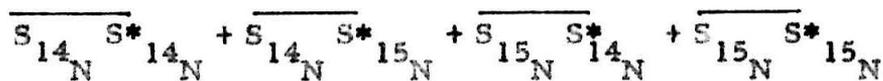
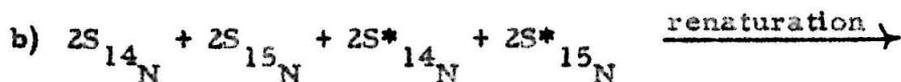
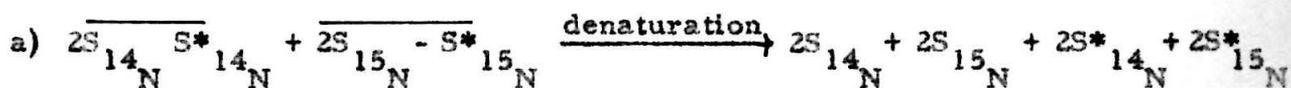
a) The subunits of hybrid  $^{14}\text{N}$ - $^{15}\text{N}$  DNA (3) may be separated in the density gradient after heat denaturation. If  $\rho_L$  and  $\rho_U$  are the densities of native  $^{15}\text{N}$  DNA and native  $^{14}\text{N}$  DNA respectively



b) When native  $^{14}\text{N}$  DNA and native  $^{15}\text{N}$  DNA are mixed, melted and renatured by heating at a temperature below the melting temperature, besides renatured  $^{14}\text{N}$  DNA and renatured  $^{15}\text{N}$  DNA a new substance (X) is formed which has a mean density intermediate between the density of renatured  $^{14}\text{N}$  DNA and that of renatured  $^{15}\text{N}$  DNA. The density heterogeneity of (X) is much greater than the density heterogeneity of renatured  $^{14}\text{N}$  DNA or that of renatured  $^{15}\text{N}$  DNA.

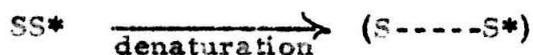


Doty believes that this material consists of renatured DNA in which one polynucleotide strand contains  $^{14}\text{N}$  and the other contains  $^{15}\text{N}$ .



Accepting for the moment Doty's assumption that the subunits of DNA are identical with the single strands of the Watson Crick helix, it is not necessary to accept Doty's interpretation of observations a) and b) above.

Observation a) that the subunits of hybrid  $^{14}\text{N}$ - $^{15}\text{N}$  DNA may be separated in the density gradient after heat denaturation, does not imply that heat denaturation alone is sufficient to separate the subunits. Intramolecular hydrogen bonds between purine and pyrimidine bases are broken during denaturation, and the Watson Crick helical structure is disorganized, but the strands could be held together by residual hydrogen bonds.



The breaking of residual hydrogen bonds resulting in strand separation could be a step subsequent to denaturation.



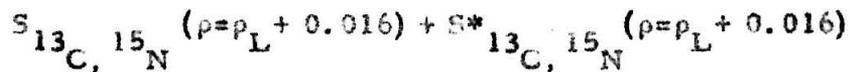
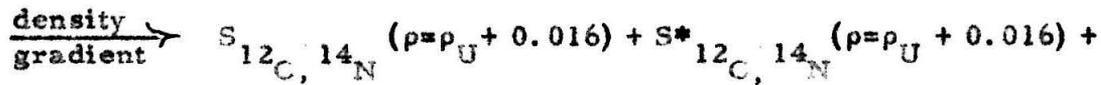
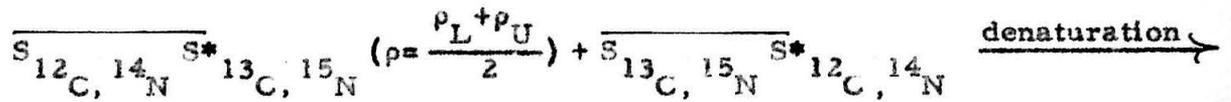
Observation b), that the substance (X) is formed when  $^{15}\text{N}$  DNA and  $^{14}\text{N}$  DNA are mixed, melted and renatured, does not imply that substance (X) consists of renatured DNA molecules each containing one  $^{15}\text{N}$  strand and one  $^{14}\text{N}$  strand. When purine and pyrimidine bases are released from intramolecular hydrogen bonding during denaturation, they form new H bonds either with other bases of the

same denatured molecule, or with bases of a different denatured molecule. During renaturation, the Watson Crick helical configuration may be regained within long segments of a denatured molecule. As they are renatured, molecules may form aggregates through inter-molecular hydrogen bonding. If  $n$  is a small integer

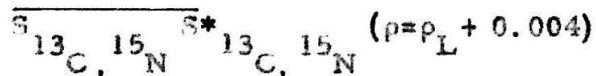
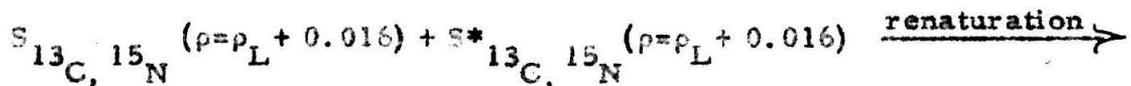


When  $^{14}\text{N}$  DNA and  $^{15}\text{N}$  DNA are mixed, melted and renatured, the aggregates which are formed would contain molecules of renatured  $^{14}\text{N}$  DNA and molecules of renatured  $^{15}\text{N}$  DNA combined in various proportions. The mean density of such an assembly of aggregates would be that observed for the material (X), and the aggregates would be expected to have a high degree of density heterogeneity, as does (X).

Thus, even with the assumption that the subunits of DNA are single strands, observations a) and b) do not provide conclusive proof that DNA molecules may be renatured after their strands have been dissociated. This question could be investigated in the following manner. DNA may be prepared in which each molecule contains one  $^{15}\text{N}$ ,  $^{13}\text{C}$  labelled subunit and one  $^{14}\text{N}$ ,  $^{12}\text{C}$  unlabelled subunit. In the density gradient after heating the DNA molecules are dissociated into labelled "heavy" and unlabelled "light" subunits. If  $\rho_L$  and  $\rho_U$  are the densities of the  $^{13}\text{C}$ ,  $^{15}\text{N}$  DNA and the  $^{12}\text{C}$ ,  $^{14}\text{N}$  DNA, respectively, we may write



Because of the large density increment provided by the  $^{13}\text{C}$  label, the heavy subunits, free from significant contamination with light subunits, may be removed from the density gradient. It would then be possible to determine if heating of the isolated heavy subunits results in 1) an increase in biological activity, 2) a decrease in buoyant density from  $\rho_1 = \rho_L + 0.016$  to  $\rho_2 = \rho_L + 0.004$ , and 3) the return of linear structures in electron micrographs.



The renaturability of the isolated heavy subunits would not prove Doty's hypothesis, since the subunits themselves may be Watson Crick helices, but the isolated heavy subunits should be renaturable if Doty's hypothesis is correct.

References

1. P. Doty, J. Marmur, J. Eigner and C. Schildkraut, PNAS 46, 461 (1960).
2. J. Marmur and D. Lane, PNAS 46, 453 (1960).
3. M. Meselson and F. W. Stahl, PNAS 44, 671 (1958).

## PROPOSITION II

A derivation of Montrol and Simha of the fragment size distribution expected in random depolymerization is considerably simplified by the application of an elementary theorem of probability theory.

Montrol and Simha (1) used the multinomial theorem to evaluate the result of introducing a number of scissions randomly into a population of polymer molecules. A simple alternative method is proposed.

Let us distribute  $K$  breaks randomly among  $N$  molecules each composed of  $n + 1$  monomers. Let  $\lambda = \frac{K}{N} = np$  be the average number of breaks per molecule, where  $p$  is the probability that a bond will be broken. Consider the fundamental probability set (FPS)<sup>2</sup> of molecules with breaks in all possible positions, in which the fragments of a given molecule are numbered.



An interchange of the  $i^{\text{th}}$  and  $j^{\text{th}}$  fragments of each molecule in the FPS is a one-one mapping of the set into itself, and hence does not change the probabilities. Thus each of the  $h + 1$  pieces has the same size distribution. The number of ways of introducing  $h - 1$  breaks into the

$n - z$  linkage remaining after a piece of size  $z$  has been broken from one end is

$$\binom{n-z}{h-1}$$

hence

$$\binom{n-z}{h-1} / \binom{n}{h}$$

is the fraction of pieces of size  $z$  among fragments of molecules with  $h$  breaks.

Since in all

$$N \binom{n}{h} \left(\frac{\lambda}{n}\right)^h \left(1 - \frac{\lambda}{n}\right)^{n-h} (1+h)$$

pieces are obtained from molecules with  $h$  breaks, the number of size  $z$  is

$$N(h+1) \binom{n-z}{h-1} \left(\frac{\lambda}{n-\lambda}\right)^h \left(\frac{n-\lambda}{n}\right)^n \quad h = 1, \dots, n-z+1$$

Summing over  $h$  gives the total number of fragments of size  $z$  resulting from the scissions.

$$N \left(\frac{n-\lambda}{n}\right)^n \sum_{h=1}^{n-z+1} (h+1) \binom{n-z}{h-1} \left(\frac{\lambda}{n-\lambda}\right)^h$$

Now this may be expressed as

$$N \left( \frac{n-\lambda}{n} \right)^n \left( \frac{\lambda}{n-\lambda} \right)^2 (n-z) \sum_{h=2}^{n-z+1} (h-1) \frac{(n-z-1)!}{(h-1)!(n-z-h+1)!} \left( \frac{\lambda}{n-\lambda} \right)^{h-2} +$$

$$2 \left( \frac{\lambda}{n-\lambda} \right) \sum_{h=1}^{n-z+1} \binom{n-z}{h-1} \left( \frac{\lambda}{n-\lambda} \right)^{h-1}$$

and setting  $h' = h - 2$   $h'' = h - 1$

$$N \left( \frac{n-\lambda}{n} \right)^n \left( \frac{\lambda}{n-\lambda} \right)^2 (n-z) \sum_{h'=0}^{n-z-1} \binom{n-z-1}{h'} \left( \frac{\lambda}{n-\lambda} \right)^{h'} + 2 \frac{\lambda}{n-\lambda} \sum_{h''=0}^{n-z} \binom{n-z}{h''} \left( \frac{\lambda}{n-\lambda} \right)^{h''} =$$

$$N \left( \frac{n-\lambda}{n} \right)^n \left( \frac{\lambda}{n-\lambda} \right)^2 (n-z) \left( 1 + \frac{\lambda}{n-\lambda} \right)^{n-z-1} + 2 \left( \frac{\lambda}{n-\lambda} \right) \left( 1 + \frac{\lambda}{n-\lambda} \right)^{n-z} =$$

$$N \left( 1 - \frac{\lambda}{n} \right)^{z+1} \left( \frac{\lambda/n}{1-(\lambda/n)} \right)^2 (n-z) + 2 \left( 1 - (\lambda/n) \right)^z \left( \frac{\lambda/n}{1-(\lambda/n)} \right) =$$

$$N p (1-p)^{z-1} [p(n-z) + 2] \quad \text{pieces of size } z$$

hence the probability  $p(z) = \frac{Np(1-p)^{z-1} [p(n-z) + 2]}{N + K} =$

$$\frac{p(1-p)^{z-1} [p(n-z) + 2]}{1 + np} \quad z \leq n$$

$$\frac{(1-p)^n}{1 + np} \quad z = n + 1$$

#### References

1. E. W. Montrol and E. Simha, J. Chem. Physics 8, 721 (1940).
2. J. Neyman, First Course in Probability and Statistics, Henry Holt (1950), p. 15.

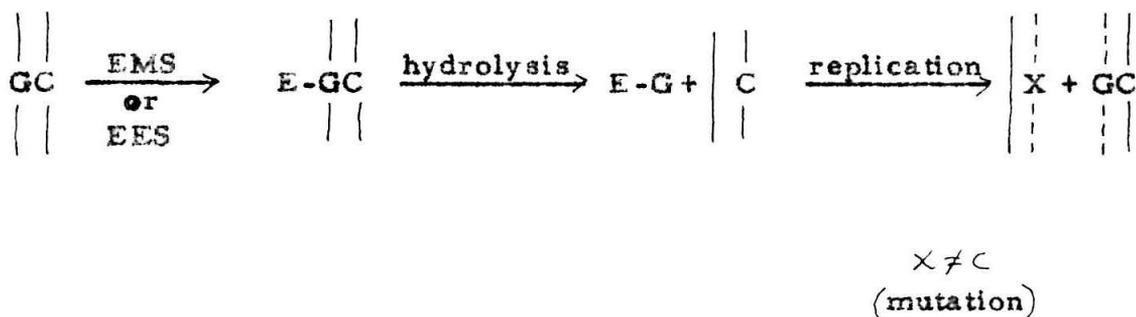
## PROPOSITION III

A mechanism for the mutagenic action of ethyl methane sulfonate is proposed.

When bacteriophage T2 or T4 are treated in vitro with ethyl methane sulfonate (EMS) and used to infect bacteria, a ten fold to one hundred fold increase in the mutation frequency is observed among their progeny (1, 2). The clone size distribution of the mutants derived from EMS treated phage indicates that EMS does not induce a mutation in the treated phage, but that the progeny of such treated phage have an increased probability of being mutant (3, 4). Assuming the Watson Crick model for replication, this probability is essentially constant for each complementary DNA strand formed by a treated template DNA strand.

Bautz and Freese (5) observed that ethyl ethane sulfonate (EES), which has a mutagenic action similar to that of EMS, is capable of selective alkylation of phage DNA at the seven position of guanine residues. They showed that the ethylated guanine residues were susceptible to loss from the DNA by hydrolysis of the N-glycoside linkage under conditions of physiological pH and temperature. Methyl methane sulfonate, which is essentially nonmutagenic under conditions in which it has been studied, also alkylates at the 7 position of guanine, but the methylated guanines are much less susceptible to hydrolytic loss from

the DNA than in the case of ethylation. On the basis of this observation, Freese proposed the following mechanism for the mutagenic actions of EMS and EES. When a 7 ethylated guanine residue is lost from the alkylated DNA, the base pairing restrictions for replication at the genetic locus in question are also lost. This allows the occurrence of a "mistake" in DNA replication, i. e., a mutation at the gene level. The postulate can be schematized as follows.



In this scheme, X stands for the base which has replaced cytosine at the indicated genetic site, and could represent any of the four bases A, G, T, C with equal probability. (If a mutation has occurred, then  $X \neq$  cytosine.)

However, the mechanism proposed by Freese does not adequately explain the following experimental results.

- 1) EMS is not capable of reverting mutants which it has induced
- (3). If Freese's explanation were correct, this would not be expected, as there should be a class of mutants such that in the above scheme

X = G. These mutants would result from base pair substitutions of the form  $GC \xrightarrow{\text{EMS}} CG$ , and since by the same mechanism we can have  $CG \xrightarrow{\text{EMS}} GC$ , they should be revertable.

2) Base analogues (5 bromo-uracil, 2 amino-purine) are capable of reverting most EMS mutants (10). Thus most EMS mutations result from transitions (substitutions of pyrimidines for pyrimidines, purines for purines) and not transversions (substitutions of pyrimidines for purines and vice versa)(6). Freese accepts this interpretation (5), but inconsistently proposes a mechanism that would product transversions twice as frequently as transitions. This can be seen from the above scheme, since for a given mutant X could be A, G or T with equal probability, but only in case X = T has a transition taken place.

The possible origin of mutations by tautomeric shifts in the DNA bases was first pointed out by Watson and Crick (7). Meselson (8) proposed that the mutagenic action of 5 bromo uracil might result from an ionization of the N1 proton, induced by the electronegative substituent bromine. More recently, Freese (6) also proposed that base analogue induced mutations in T4 result from improper base pairing. Subsequent to the initial writing of this proposition, Green and Krieg suggested that EMS mutagenesis might be explained in terms of a tautomeric shift in ethylated guanine (4).

The following mechanism for EMS mutagenesis is proposed as an alternative to that of Freese.

The effect of alkylation at the 7 position on desoxyguanosine in phage DNA is to place a formal positive charge on the imidazole ring nitrogens.

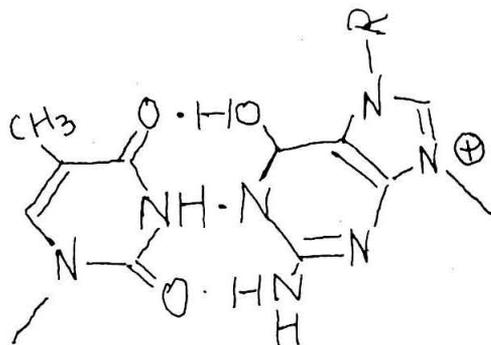


The resulting increase in electronegativity leads to a decrease in  $\pi$  electron density in the pyrimidine ring and the N1 proton becomes more acidic. The magnitude of increase of the  $pK_a$  should be comparable to that produced by a quaternary ammonium substituent in pyrrole, about 2 pK units. In addition to the pK effect, there is probably also an effect on the amide-enol tautomeric ratio. This effect is analogous to one observed by Mason in some N heteroaromatic hydroxy compounds, in which a lowering of ring  $\pi$  electron density strongly favored the enol tautomer (9).\*

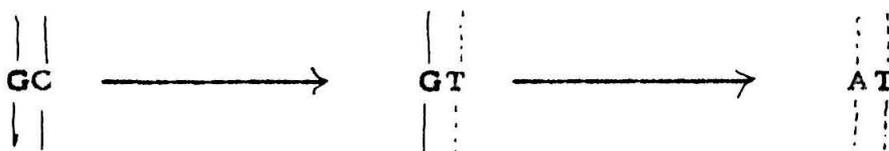
During DNA replication, the ethylated guanine residue may lose the N1 proton or may undergo a tautomeric shift to the enol form. Under these circumstances, the guanine residue will pair with thymine rather than cytosine.

---

\*While these effects expected for 7 alkylation of a desoxyguanosine monomer might be somewhat modified for desoxyguanosine in the DNA polymer, it is unlikely that they would be completely reversed.



The mutation which is produced in the progeny molecule after one more round of replication consists of a substitution of the AT pair for the GC pair present in the corresponding position in the EMS treated parental molecule.



To distinguish between these two hypotheses, the following experiments could be carried out.

An attempt could be made to synthesize a GC polymer containing 7 ethyl guanine, using the Kornberg polymerase system. The titration curve of such a polymer could be compared with that of the normal GC polymer, to determine the effect of 7 ethylation on the  $pK_a$  of the N1 proton of guanine.

The "ordinary" GC polymer could be ethylated with EMS and used as a primer in the polymerase system. The following could be

determined.

1. The extent of ethylation of guanine residues.
2. The extent of deguanylation of the polymer.
3. The amount of  $^{32}\text{P}$  labelled deoxyadenosine triphosphate incorporated into newly formed GC polymer, when the ethylated primer is used.
4. The amount of tritium labelled thymidine triphosphate incorporated simultaneously.
5. The effect of pH on 2, 3, and 4.

If the Freese hypothesis is correct, the incorporation of thymine and adenine should depend only on the extent of deguanylation, and should be enhanced by low pH, which facilitates the hydrolysis.

If the alternative hypothesis is correct, the incorporation of adenine should be negligible. The incorporation of thymine should depend upon the total number of ethylated guanine residues remaining in the polymer, rather than upon the extent of deguanylation. The incorporation may be enhanced by high pH, which facilitates the ionization of the N1 proton of ethylated guanine.

References

1. A. Loveless, Nature (1958), 181, 1212-13.
2. A. Loveless, Proc. Roy. Soc. (London) (1959), Ser. B, 150, 497-508.
3. D. Green, Cold Spring Harbor Meeting, 1959 Phage Information Service #14, pp. 1-2.
4. D. M. Green and D. R. Krieg, Proc. Nat. Acad. Sci. (1961), 47, 64-71.
5. E. Bautz and E. Freese, Proc. Nat. Acad. Sci. (1960), 46, 1585-93.
6. E. Freese, Proc. Nat. Acad. Sci. (1959), 45, 622-633.
7. J. D. Watson and F. H. C. Crick, Cold Spring Harbor Symp. Quant. Biol. (1953), 18, 123-131.
8. M. Meselson, personal communication.
9. J. Mason, J. Chem. Soc. (London) (1957), 5010-5019.
10. R. Sinsheimer, personal communication.

## PROPOSITION IV

An experiment is proposed to test a theory of the mechanism of the formation of Segment Long Spacing collagen.

Tropocollagen macromolecules are stiff rods about 2800 Å in length, with a diameter of 15 Å (1, 2). They are composed of three polypeptide chains in a helical configuration (3). Native collagen consists of an ordered array of protofibrils. Each protofibril is a linear polymer of tropocollagen molecules associated end to end and oriented in a common direction. The protofibrils are associated side by side and staggered with respect to each other at intervals of  $n/4 \times$  the molecular length. Thus the tropocollagen molecules in native collagen are said to be in "heteroregister" or in "quarter stagger" arrangement (4). Tropocollagen macromolecules in very dilute acetic acid form the native collagen structure when dialysed against 1% NaCl. If instead a nucleoside triphosphate, such as ATP, is added to the solution, a different type of structure, the Segment Long Spacing (SLS), is formed. The SLS "monomer" is the result of side by side association of tropocollagen molecules, forming a crystallite of length equal to the molecular length. The molecules in the SLS structures are said to be "in homoregister" or "unstaggered". When SLS structures are formed from protofibrils, they are called SLS polymers.

It has been suggested (5) that the negatively charged ATP molecules allow the formation of the SLS structure because they interact with the regions of positive charge on the tropocollagen macromolecule.

which prevent the formation of SLS. The "neutralization" of the positive charge alters the "interaction profile" of the tropocollagen molecule, and the SLS structure is formed. The moles of ATP bound in the formation of SLS have been found to be equivalent to the moles of lysine and hydroxylysine in the collagen sample (6). Thus the epsilon amino groups of collagen lysine and hydroxylysine residues may carry the positive charges which prevent the formation of SLS. If this hypothesis is correct, it might be possible to form SLS structures in the absence of ATP, by chemically modifying the free amino groups of collagen.

Tropocollagen is soluble at pH 5 and low ionic strength, and it is probably correct to assume that the excess of positive over negative charges is not great at this pH. Examination of the amino acid composition of native collagen (7) reveals that of 419 amino acid residues, 37 are dibasic (20 arginines, 15 lysines and hydroxylysines, and 2 histidines) and 49 are dicarboxylic (glutamics and aspartics). Of the 49 carboxyl side chains, 18 are present as amides.\* The free carboxyl groups of collagen have an average  $pK_a$  of 3.5 and at pH 5 are essentially all ionized (7), while the  $pK_a$ 's for arginine and lysine side chains are 14 and 11. Thus at pH 5 the net positive charge on the molecule could be removed by chemical modification of the lysine and hydroxylysine residues.

---

\*Some preparations of collagen have less amide nitrogen, and the isoelectric point, normally at pH 7.5, may be as low as pH 5.5.

Comparison of electron micrographs of SLS stained by cationic uranium and by phosphotungstic acid, which is specific for arginine, residues under these conditions, suggests that all polar areas of the molecule contain both acidic and basic groups (4). Thus it is unlikely that there would be regions of high positive charge on the molecule when the net charge is zero or negative.

The basic side chains of tropocollagen may be selectively modified by alkylation, by acylation, or by deamination (8, 9, 10, 11). The reactions of aldehydes with the basic groups would not be useful in the present study as they would lead to "tanning" of the ichthycol solutions by cross-linking of molecules.

Among the alkylating agents used to modify proteins, dinitrofluorobenzene (DNFB) is one of the most selective for basic groups, reacting with epsilon and N terminal alpha amino groups, and with imidazole groups in proteins. In contrast to the majority of the other alkylating agents, it does not modify the free carboxyl side chains of proteins. Collagen has been treated with DNFB (11, 12), but at least 20% of the epsilon amino groups appear to be inaccessible to the reagent. Thus DNFB does not appear to be the reagent of choice for the selective modification of lysine residues.

The acylating agents as a class are more selective in protein modification than are the alkylating agents because they do not readily react with free carboxyl groups. A variety of agents have been used

for selective acylation, but most of them do not react quantitatively with the free amino groups of proteins. Even ketene, which has been considered one of the best protein acylating agents, does not quantitatively modify the free amino groups of all proteins. However, a procedure of choice for acylation of collagen has been reported by Green et al. (13). They were able to introduce a quantity of N acetyl groups into collagen equivalent in moles to the content of lysine and hydroxylysine residues, using acetic anhydride in sodium acetate at pH 8. Since guanidino groups are not acetylated under these conditions, they concluded that the free amino groups of collagen had been totally acetylated.

As a supplement to the acylation procedure discussed above, deamination offers some advantage in the study of the effect of modification of the basic groups of collagen on the formation of SLS. It is capable of modifying both the majority of the free amino groups and a substantial fraction of the guanidino groups of the arginine residues.

Treatment of collagen with  $\text{NaNO}_2$  in acetic acid results in the quantitative removal of the epsilon amino groups of lysine and hydroxylysine, and converts 20% of the guanidino groups of the arginine residues to cyanamide groups (14). If the reaction is carried out in strong mineral acids at pH 2, up to 2/3 of the arginines may be modified (11).

To study the ability of the modified tropocollagen to form SLS in the absence of nucleoside triphosphates, the following experiments could be done.

1. Tropocollagen in acetic acid solution could be treated with  $M/16 \text{ NaNO}_2$  at  $38^\circ\text{C}$  (15). After dialysis against water, the resulting solution or gel could be examined in the electron microscope for the presence of SLS.

2. Tropocollagen in sodium acetate solution at pH 8, ionic strength  $> 0.4$  could be acetylated with acetic anhydride in the cold. To determine whether SLS can be formed without addition of ATP, the modified collagen could first be dialyzed against citrate buffer at pH 4, ionic strength 0.2, and then against water.

References

1. H. Boedtker and P. Doty, J. Am. Chem. Soc. (1956), 78, 4267-4280.
2. C. E. Hall and P. Doty, J. Am. Chem. Soc. (1958), 80, 1269.
3. A. Rich and F. H. C. Crick, Nature (1955), 174, 915-916.
4. A. J. Hodge and F. Schmitt, Proc. Nat. Acad. Sci. (1960), 46, 186-197.
5. A. J. Hodge, 4th Int. Conf. Elect. Microscopy (1960), pp. 119-139.
6. J. Cross, H. Highberger and F. O. Schmitt, Proc. Nat. Acad. Sci. (1954), 40, 679-688.
7. J. H. Bowes and R. H. Kenten, Biochem. J. (1948), 43, 358-365.
8. R. M. Herriot, "Reactions of Native Proteins with Chemical Reagents," in Advances in Protein Chemistry, Vol. III, (1947), pp. 169-225.
9. F. W. Putnam, "Chemical Modifications of Proteins," in Neurath and Bailey, The Proteins, Vol. 1, Pt. B (1953), Academic Press, pp. 893-972.
10. H. Fraenkel-Conrat, "Chemical Reactions of Proteins," in Greenberg, Amino Acids and Proteins, Thomas (1951), pp. 532-585.
11. K. H. Gustavson, "The Chemistry and Reactivity of Collagen," Academic Press (1956), pp. 246-257.
12. J. H. Bowes and J. A. Moss, Nature (1951), 168, 514-515.
13. R. W. Green, K. P. Ang and L. C. Laue, Biochem. J. (1953), 54, 181-187.
14. J. H. Bowes and R. H. Kenten, Biochem. J. (1949), 44, 142-152.
15. J. St. L. Philipot and P. A. Small, Biochem. J. (1938), 32, 543-551.

## PROPOSITION V

An investigation of the crosslinking in the  $\gamma$  component of collagen is proposed.

When tropocollagen is thermally denatured, the ordered helical structure of the molecules is lost. The bulk of the material is converted into two new substances, components  $\alpha$  and  $\beta$ , which are distinguishable by their amino acid compositions, and their chromatographic and sedimentation behavior. The molecular weights of the  $\beta$  and  $\alpha$  components of denatured collagen are  $2/3$  and  $1/3$  of the molecular weight of the parent tropocollagen, and they are produced in equal molar quantities. It is thought that the  $\alpha$  component represents two of the three polypeptide chains of collagen, the  $\beta$  component being the third chain which has separated from the other two. About 5% of the tropocollagen is converted to a third or  $\gamma$  component by denaturation (1). The  $\gamma$  component has sedimentation properties distinct from the  $\alpha$  and  $\beta$  components, and has the unique property of renaturability. Thus  $\gamma$  collagen may be reversibly denatured, and after renaturation can form native collagen fibril or SLS structures. It has been proposed that some type of crosslinking between the three polypeptide chains of  $\gamma$  collagen gives it the property of renaturability. In the majority of the molecules of a collagen sample, the crosslinking is limited to the two chains of collagen which

make up the  $\alpha$  component, and denaturation is not reversible as it results in the loss of the third polypeptide chain.

The following experiment is suggested to investigate the extent and nature of the crosslinking between the three chains of  $\gamma$  collagen. It is possible to break collagen into half and quarter molecules by sonic irradiation (2). The collagen fragments associate selectively into half and quarter SLS structures when ATP is added to the solution. The renaturability of sonic fragments of  $\gamma$  components may be investigated by determining the proportion of the sonicated sample which can form SLS type structures after denaturation. If sonication breaks the crosslinks responsible for renaturability, or if there are only one or two crosslinks per molecule, the amount of SLS recovered should decrease with increasing sonication. If the crosslinks are resistant to sonication and are present in large numbers along the length of the molecule, then most of the denatured sonic fragments should be recoverable as SLS.

#### References

1. A. J. Hodge, to be published.
2. A. J. Hodge and F. O. Schmitt, PNAS, 44, 418 (1958).