

The Arrangement of Information in DNA Molecules

CHARLES A. THOMAS, JR.

From the Department of Biophysics, The Johns Hopkins University, Baltimore

ABSTRACT The anatomy of DNA molecules isolated from mature bacteriophage is reviewed. These molecules are linear, duplex DNA consisting mainly of uninterrupted polynucleotide chains. Certain phage (T5 and PB) contain four specifically located interruptions. While the nucleotide sequence of most of these molecules is unique (T5, T3, T7, λ), some are circular permutations of each other (T2, T4, P22). Partial degradation of these DNA molecules by exonuclease III predisposes some of them to form circles upon annealing, but indicating they are terminally redundant.

At the present time, the only available DNA molecules that contain discrete information are those liberated from viruses. In the case of bacterial viruses, where the subject is most highly developed, it is found that most species of bacteriophage liberate a single duplex DNA molecule. The exceptions are ϕ x or M13 phage which contain a circular, single polynucleotide chain. This DNA molecule has been properly identified as the viral chromosome, or that physical structure that is responsible for the genetic map. Since bacteriophage chromosomes are the products of replication and recombination, a study of their anatomy contributes in a fundamental way to understanding these events. In view of the fact that bacteriophage DNA molecules can become incorporated into the DNA of certain hosts to produce lysogenic bacteria, a study of the virus DNA will undoubtedly lead to a greater knowledge of the anatomy of the chromosomes of bacteria, and ultimately of the chromosomes of higher cells. Thus the study of the genetic and physical structure of phage DNA molecules represents our first efforts in what will inevitably become the larger subject to *molecular cytogenetics*, the objectives of which would be not only to relate the genetic map with the genetic molecules, but also to understand the structural basis for the controlled transcription of the DNA molecule.

The Molecular Weight and Molecular Length of Bacteriophage DNA Molecules

The great length of intact DNA molecules has led to unexpected difficulties in determining molecular weight and length. These difficulties have justifiably received considerable attention in recent years and I will not contribute to this ongoing debate here. As might be expected, however, at the point

at which physical chemical procedures become increasingly difficult to apply, cytological procedures become useful.

In this regard the invention of Kleinschmidt, Lang, and Zahn (1) has played an important role. In these experiments, purified DNA molecules

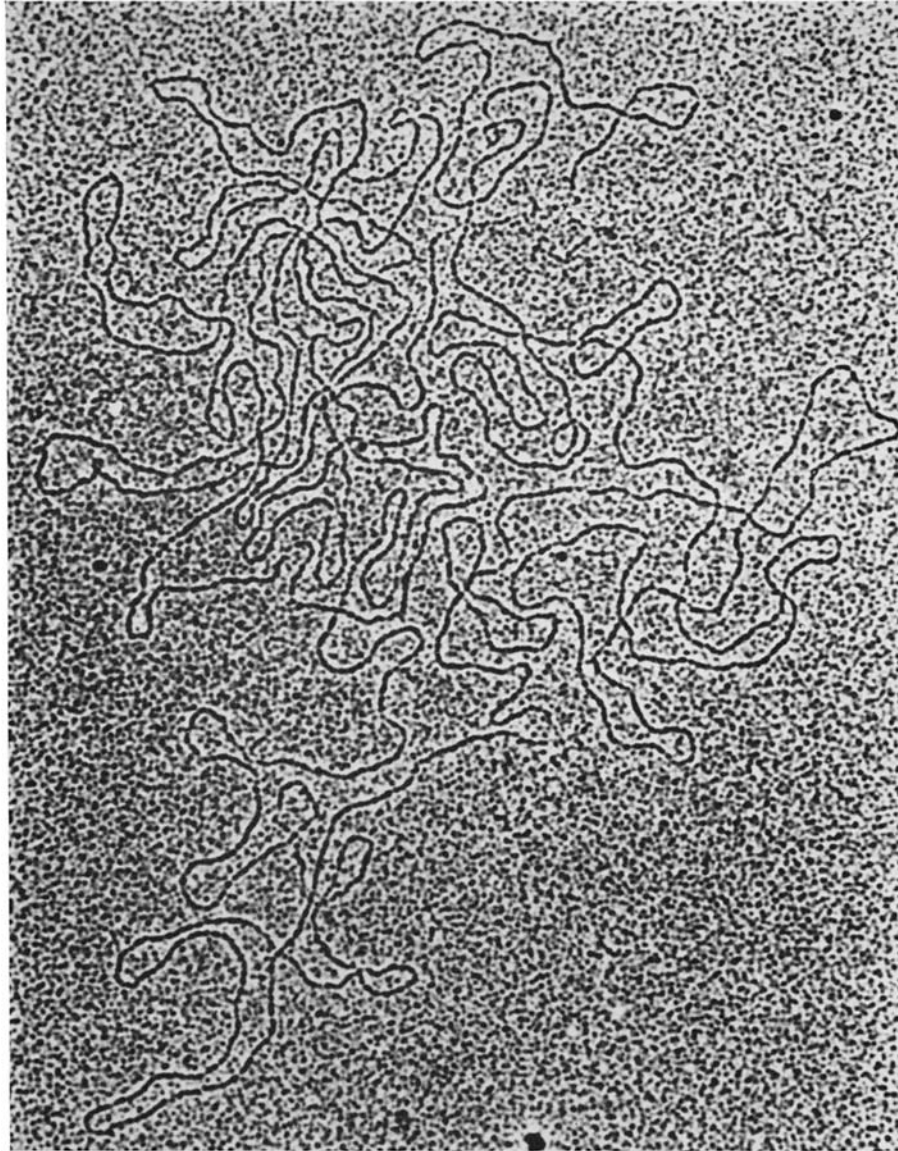


FIGURE 1. A typical T2 DNA molecule as visualized by protein film technique. The more compact configuration is a result of the presence of 3.6 M NaClO_4 in DNA-protein mixture. This compaction makes it easier to photograph very long molecules. Unfortunately it produces a slight (5%) contraction in mean length. $\times 47,000$.

are mixed with any basic globular protein and the complex is spread on a clean air-water interface. After a brief "aging" period, the surface is touched with an electron microscope grid bearing a thin carbon film. The resulting grid is shadowed, often from all directions and then examined at low magnification in the electron microscope. A typical result showing the thickened DNA molecule can be seen in Fig. 1.

The contour length of the fiber can be measured easily. However, before this length can be related to molecular weight two other pieces of information

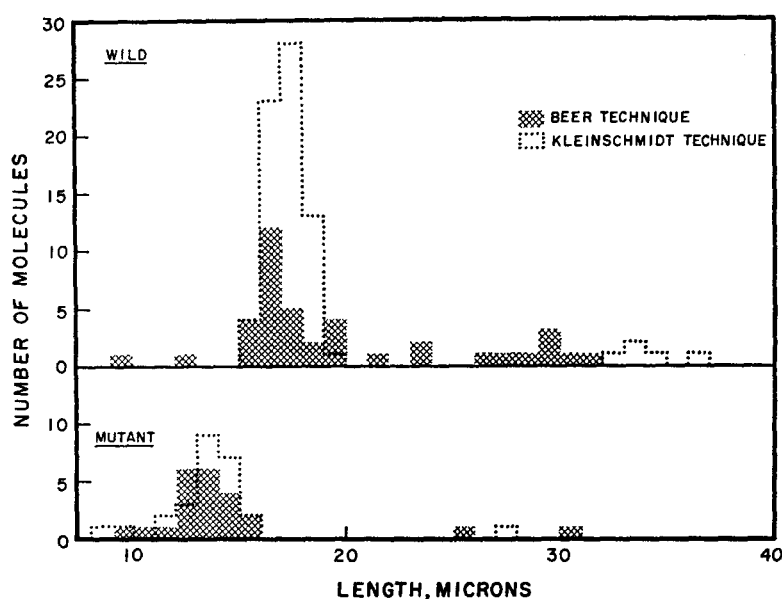


FIGURE 2. The length of λ -DNA molecules. These histograms show the results of measuring $\lambda c + +$ and $\lambda c b_2 b_5$ DNA molecules on smooth carbon grids (Beer technique) or in protein films (Kleinschmidt technique). Both linear and circular molecules have the same contour length.

are required: the over-all magnification and the linear density of the DNA molecule on the electron microscope grid. The magnification problem can be partially solved by intercalating electron micrographs of grating replicas, the spacing of which can be independently checked. Unfortunately, this does not solve the problem completely because the magnification varies slightly with position on the plate, and wanders during a series of micrographs.

The determination of the linear density of the DNA molecule on the electron microscope grid is a more serious problem. Even granting that the crystallographic B form of the DNA duplex has a linear density of 196 daltons/A (2), there is every reason to expect that the DNA-protein complex will be extended or shrunk by the time it dries on the electron microscope grid. To check this requires micrographs of molecules of known molecular weight. The

only method available for measuring the molecular weight of large DNA molecules is ^{32}P autoradiography or "star counting," which, though absolute in principle, rests on its own assumptions (3). A few years ago we measured the molecular weight of T2 DNA at 130 million (4). By measuring the relative sedimentation rate of λ -DNA molecules and T2 DNA molecules, and applying an empirical relation between relative S and relative M obtained by sedimenting whole and half molecules of T2 through sucrose gradients (5), Burgi and Hershey assigned at molecular weight of 31×10^6 to λ -DNA. Using this DNA molecule, Dr. Lorne MacHattie and I attempted to identify the correctness of the linear density of DNA on electron microscope grids. We found that the measured length of λ -DNA molecules was the same (17.2 μ) either in protein films or when streaked on smooth carbon-covered grids without protein. A deletion mutation which removes 23% of the DNA (as determined from the altered density of the phage particle) was found to liberate DNA molecules which are 13.2 μ or 23% shorter (7). Thus the observed length of λ -DNA was exactly that predicted for a B-form molecule with a molecular weight of 33 million and a linear density of 196 daltons/A. Our measurements are shown in Fig. 2. As part of this study, we were able to demonstrate that the rapidly sedimenting form, discovered by Hershey, was circular.

Confirmation of this value for the linear density has also come from other sources: the contour length of ϕ x replicative form (8), recent length and autoradiographic determinations on T7 DNA (9), and our more recent length measurements on T2 and T*2 (nonglucosylated) DNA molecules. These results together with a compilation of molecular lengths measured in our laboratory by Dr. MacHattie are shown in Fig. 3. On the ordinate is plotted molecular weight (of the polymeric sodium salt); on the abscissa is plotted molecular length. The straight line is drawn for 196 daltons/A. The circled points correspond to molecules whose molecular weights are derived directly or indirectly from ^{32}P autoradiography.

The conclusion is that all these phage chromosomes are linear duplex DNA molecules and that molecular length appears to be a good measure of molecular weight. Lengths and calculated molecular weights are compiled in Table I.

Certain problems remain. The molecular weight of T2 DNA is 10% too high or the lengths of the molecules are too short by the same amount. This is not the result of glucose as can be seen by the identity of the histogram for T2 and T*2 DNA molecules.

Heterogeneity in Measured Lengths

There is the annoying problem of the heterogeneity in the observed lengths. This heterogeneity, which amounts to about $\pm 5\%$, is trivial by the standards of a polymer chemist, but it is enormous by the standards of a geneticist.

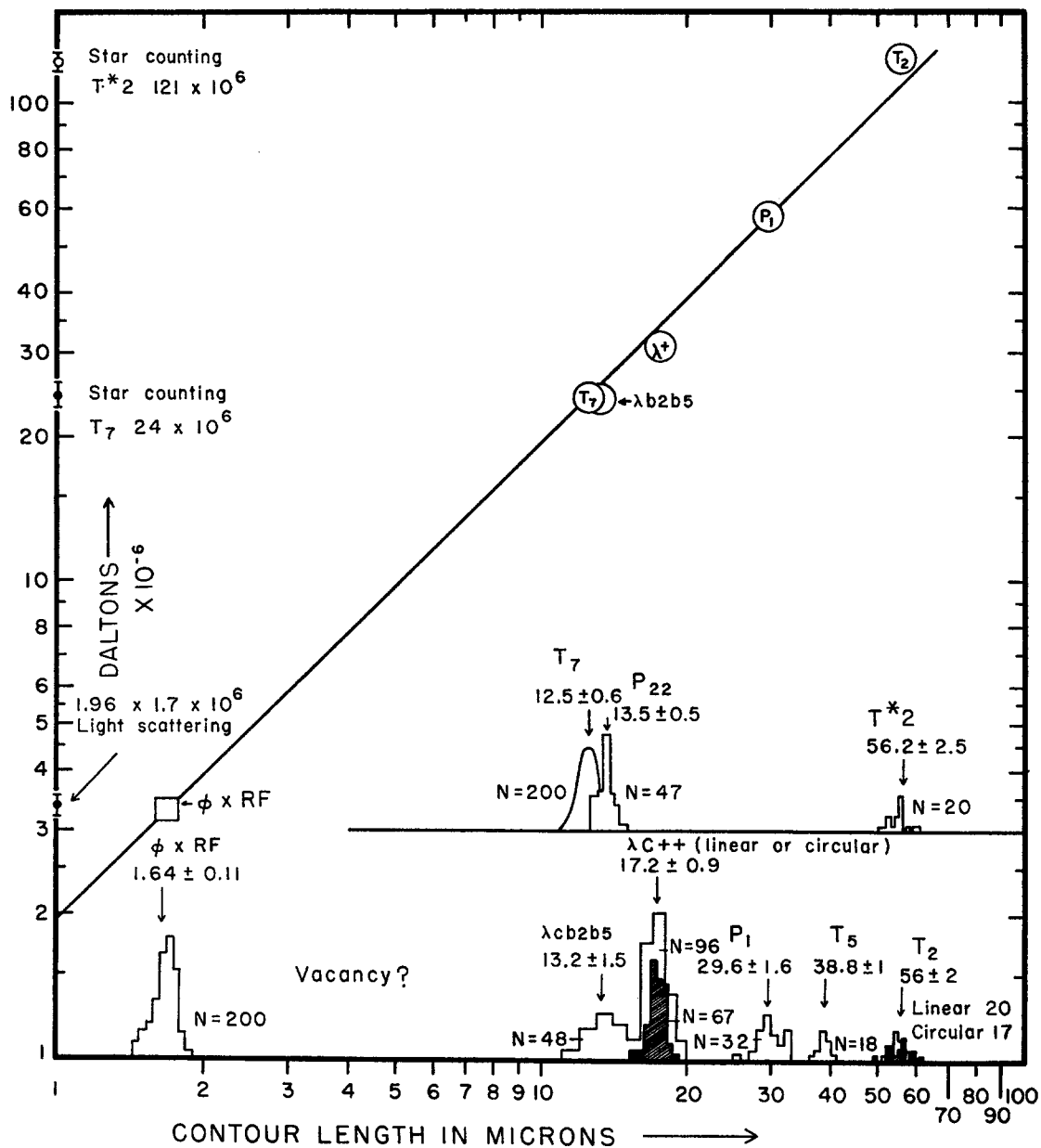


FIGURE 3. The molecular weight and length of phage DNA molecules. On the ordinate is plotted the molecular weight of phage DNA molecules; on the abscissa the molecular length. The histograms of observed molecular length are superimposed. The straight line is drawn for a density of 196 daltons/A corresponding to the B form of DNA. Most of the molecular lengths shown have been measured in this laboratory. The autoradiographic determination of the molecular weight of T7 DNA by Rubenstein and Leighton (9) along with Hans Bode's (Yale) molecular length of $12.5 \pm 0.6 \mu$ (which we have confirmed) provides an important check on the observed linear density. The molecular weight of P1 DNA has been estimated at 58 million by Tomizawa and Anraku (17) by sedimentation rate relative to T2 DNA and we have confirmed their value.

Part, but not all, of this variance in length can be attributed to variability in magnification. Part can be attributed to stretching of the molecules, since those molecules which assume extended configurations in sheared films are found to be slightly longer. Finally, in the case of some linear molecules, there may be significant variability in nucleotide content.

For DNA molecules containing identical numbers of nucleotides, the lower limit to the heterogeneity in length should depend on thermal motion of the molecule. By applying the Boltzmann theorem one may calculate that

TABLE I
LENGTH AND CALCULATED MOLECULAR
WEIGHTS OF VARIOUS PHAGE DNA'S

Phage	\bar{L}	σ	σ/\bar{L}	M_L	M		Reference
			<i>per cent</i>				
T2	56	2	3.6	118	130	(Stars)	4
T*2	56.2	2.5	4.4	110	121		
T5	38.8	1	2.6	76			
P1	29.6	1.6	6.0	58	58	(Sucrose gradient)	17, 18
λ c++	17.2	0.9	5.2	33.7	31	(Sucrose gradient)	7, 6
λ cb ₂ b ₆	13.2	1.5	11.0	25.9	25	(Sucrose gradient)	7
P22	13.5	0.5	3.7	26.5			
T7	12.5	0.6	4.8	24.5	24	(Stars)	9, 18
ϕ xRF	1.64	0.11	6.7	3.2	3.4	(2 \times l.s.)*	8

The mean length \bar{L} and the standard deviation for a single length measurement σ are listed in the second and third columns. The molecular weight calculated from the B form linear density ($M_L = 1.96 \times 10^6 \cdot \bar{L}$) is to be compared with M determined directly or indirectly by autoradiography.

* This is twice the molecular weight determined from the light scattering. See reference 8.

this variance in length (σ_L^2) is given by the expression

$$\frac{\sigma_L^2}{\bar{L}} = \frac{kT}{EA} \quad (1)$$

where k is the Boltzmann constant, T the absolute temperature, E is Young's modulus, and A the cross-sectional area of the molecule. It is not easy to calculate or measure the Young modulus of a single molecule like DNA, but $E \cdot A$ has been estimated at 3.5×10^{-4} dynes for helical polypeptides (10). In order to be conservative we assume that DNA is 10 times "softer" than a helical polypeptide ($EA = 3.5 \times 10^{-5}$ dynes). When substituted into equation 1 this gives $\sigma_L^2/\bar{L} = 10^{-9}$. For a DNA molecule 0.010 cm long this gives $\sigma_L = 3 \times 10^{-6}$ cm or 300 Å, which is much smaller than the observed variance. The histograms shown in Fig. 3 are at least 100 times broader than their thermal limit. Thus there is no reason *in principle* that this procedure cannot be further refined and made more precise.

What is greatly needed is a rod-like particle of accurately known length, about 1 to 2 μ long, that can be used for an internal magnification reference.

On the Continuity of Polynucleotide Chains

When the DNA molecule is placed in a solvent in which the secondary structure is not stable, the component single chains should separate from each other. If the chains are continuous, two full length single chains should be produced. If interruptions exist or if there are special linkages which are broken by the denaturing treatments, then more than two chains would be liberated. This problem has a long history, dating back before the advent of the Watson-Crick model for DNA. Today, the interest in this problem turns on the question of whether functionally important special points exist along the bacteriophage DNA molecule.

The first evidence on this question came from the determination of molecular weights of intact and denatured T2/T4 DNA molecules by equilibrium band widths in CsCl density gradients (11). It appeared that most of the molecules were composed of intact polynucleotide chains, or at least they did not contain gaps in the middle third of the molecule. This subject was further studied (12) and summarized a few years ago (13).

Banding in CsCl is not the most desirable procedure for this task; it does not fractionate the single chains with respect to their length, and second, unless the two complementary chains have effectively the same density the band-widths will be too broad. The method measures the minimum molecular weight or the maximum number of gaps. Therefore, we have spent a good deal of time trying to fractionate single polynucleotide chains (16). Eventually, Studier (6) had to show us how to do it using the band sedimentation method invented by Vinograd and his collaborators (14). Using his solvents (0.90 M NaCl, 0.10 M NaOH) we still observed very broad zones when sedimentation was performed in the preparative centrifuge using labeled DNA's. Marcus Rhoades, in our laboratory, finally demonstrated this to be the result of radiochemical damage to the phage DNA molecule during growth and storage. When phage were grown for less than 1 hr in a medium containing less than 10 μ C/ml of ^{32}P or ^{14}C , very tight zones were seen when the DNA was sedimented through 5 to 20% sucrose gradients containing 0.90 M NaCl, 0.10 M NaOH. We have found it impossible to prepare intact polynucleotide chains labeled with tritium.

STANDARD COLLECTION When lightly labeled, purified T2, T*2, P1, P22, λ , or T7 phage or phenol-extracted DNA molecules are denatured by dilution into 0.20 M Na_3PO_4 and sedimented through alkaline sucrose, a single zone is found. An illustration is seen in Fig. 4 which shows the different sedimentation rates of single polynucleotide chains from T*2, P1, and T7. The distance travelled by a macromolecule through a sucrose density gradient

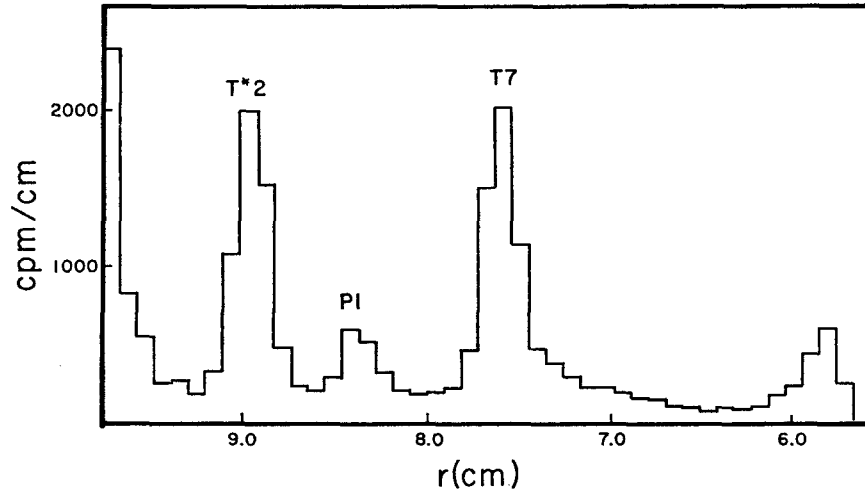


FIGURE 4. Sedimentation profile of a mixture of T*2, P1kc, and T7 DNA in an alkaline sucrose gradient.

depends on the way the density and viscosity increase with the radial distance from the axis of rotation. The appropriate measurements have been made for this solvent and the sedimentation equation has been numerically integrated (15). It turns out that the sedimentation coefficient, $s_{20,w}$, is accurately proportional to the distance the zone travels from the meniscus. The measurement of comparative S is still most accurate. Therefore John Abelson, in our

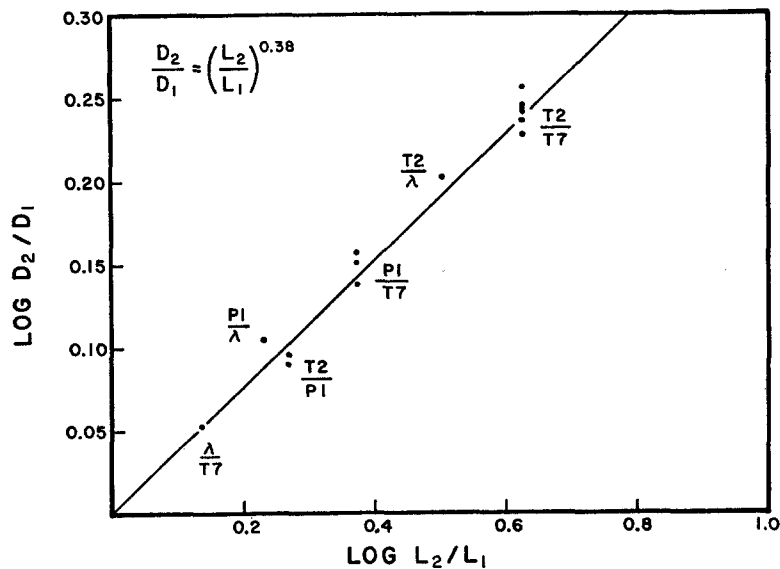


FIGURE 5. Calibration scale. The ratio of the distances traveled by a pair of denatured DNA's is plotted against the ratio of lengths of the duplex DNA molecules as determined in the electron microscope.

laboratory, sedimented the single chains derived from pairs of DNA's and measured their relative sedimentation rates by the ratio of the distance that each zone travelled. This ratio is plotted in Fig. 5 against the ratio of the molecular lengths of the duplex molecule as seen in electron micrographs and quoted in Table I. The points fall on a straight line given by

$$\frac{D_2}{D_1} = \frac{S_2}{S_1} = \left[\frac{L_2}{L_1} \right]^{0.38} \quad (2)$$

We interpret this to mean that each of these phage DNA molecules denatures to produce two continuous polynucleotide chains. The only other alternative

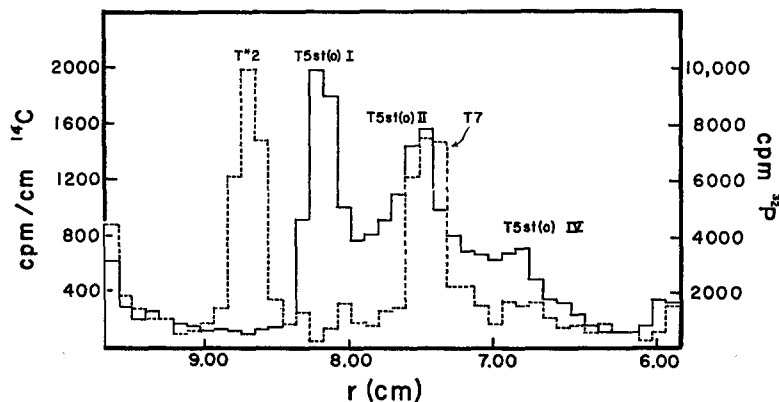


FIGURE 6. Alkaline sedimentation profiles of single polynucleotide chains from T*2, T5, and T7 DNA molecules. ^{14}C -labeled T*2 and T7 DNA's were mixed with ^{32}P -labeled T5 DNA, denatured, and sedimented through an alkaline sucrose gradient. The relative sedimentation rates of the various peaks can be measured by the ratio of the distances traveled.

we can imagine is that all the phage DNA's in this collection contain the *same* number of gaps which are so located as to produce single polynucleotide chains of indistinguishable length. This artificial objection can be eliminated by a number of experiments, the most direct being the demonstration that prior shear breakage of the duplex molecule causes a commensurate decrease in the molecular weight of the single chains derived from the fragments. This is the case with T2, T4 (12, 16, 17), and λdg (6).

The data plotted in Fig. 5 provide one with an empirical scale that can be used to measure the length of a polynucleotide chain of unknown length. This empirical scale was constructed for the purpose of measuring the length of the polynucleotide chains liberated by denaturing T5 DNA (18).

The Anatomy of the T5 DNA Molecule

When purified T5st(0) bacteriophage (or isolated DNA) is mixed with 0.20 M Na_3PO_4 and sedimented through alkaline sucrose gradients, one does not

find a single zone, as is the case with the *standard phage*, but rather four distinct zones. To illustrate these results, Fig. 6 shows the profiles found when a mixture of T*2 and T7 phage (labeled with ^{14}C thymidine) and T5st(0) (labeled with ^{32}P) is alkali-denatured and sedimented through alkaline sucrose gradients. These components can also be identified by sedimentation of alkali-denatured DNA through neutral gradients. This fact makes it possible to

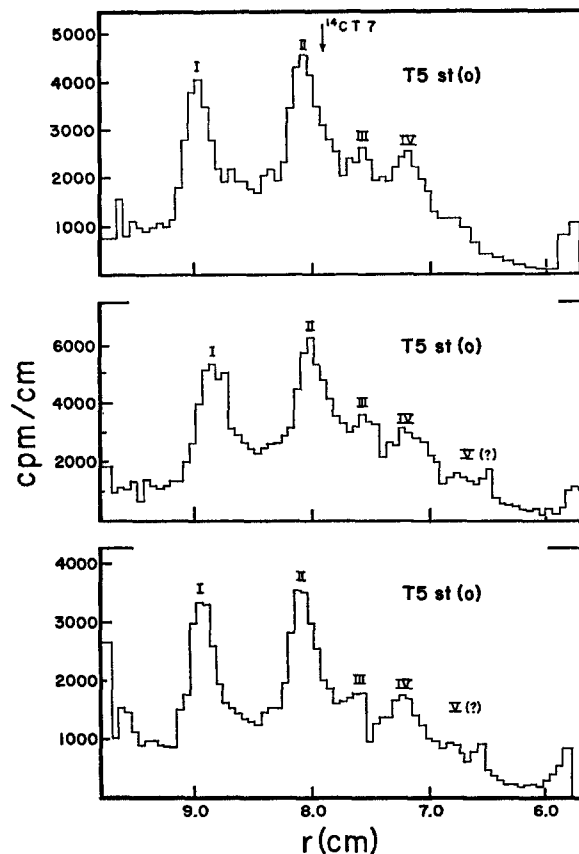


FIGURE 7. High resolution sedimentation profiles of T5st(0) in alkaline sucrose.

show that treatment with increasing pH produces no change until pH 11.89 is reached, at which point the molecule is denatured. Pure formamide produces essentially the same effect as high pH. Therefore we think that the simplest explanation is that the T5 DNA molecule contains gaps at specific locations in the molecule.

In order to quantify these observations, higher resolution alkaline sedimentation experiments were performed containing one or both of the fiducial phage (T*2 or T7) single chains. The ratio of the distance travelled by components I, II, III, and IV and the standard polynucleotide chains was meas-

ured and the ratio of their lengths calculated from the graph shown in Fig. 5. Taking the length of either T7 or T2 as known, the actual lengths of the unknown polynucleotide chains in the four peaks may be determined. The reproducibility of these sedimentation profiles may be seen in Fig. 7.

By dividing the amount of material found in these four peaks by the length of the chain determined by relative sedimentation rate, one may calculate the number of chains of each length. The results for T5st(0) are shown in Table II. There is *one* 34.0 μ and *one* 8.0 μ chain, and *two* 14.4 μ and 4.0 μ chains. When the total length of polynucleotide chain in the T5 molecule is

TABLE II
LENGTH OF POLYNUCLEOTIDE CHAINS IN T5st(0)

Peak	\bar{L}	σ	N	Experiment		Nominal	Total length
				A	B		
I	34.0	2.0	9	(1.00)	(1.00)	1	34.0
II	14.4	0.5	8	2.16	2.17	2	28.8
III	8.0	0.4	5	1.18	1.23	1	8.0
IV	4.0	0.2	8	2.13	2.44	2	8.0
V?	1.75	—	2	—	—	—	—
Total						6	78.8 μ
$2 \times 38.8 =$							77.6 μ

The average length (in microns) \bar{L} is the average of N determinations with a standard error of σ for a single determination. Experiments A and B listed refer to two typical experiments giving the counts per minute per micron for each component, normalized to one 34.0 μ chain per molecule. The nominal number of chains is shown in the next column. The final column gives the total length of polynucleotide chain which is equal to twice the observed length of the duplex T5 DNA molecule.

added up, it comes to 78.8 μ , which agreeably enough, is just twice the length of the duplex T5 DNA molecule as seen in the electron microscope! We take this agreement as an *independent* confirmation of the length and number assignments shown in Table II.

Biological Control of Interruptions in the T5 DNA Molecule

Probably the first question that comes to mind is whether the location of these gaps is under the control of the host cell or of the T5 phage. When T5st(0) is grown on four different *coli* hosts, the progeny phage produces the same profiles as shown in Fig. 7. On the other hand, when different T5 mutants are examined, different alkaline sedimentation profiles are seen (Fig. 8). We have examined five different phage in all and measured the length and number of polynucleotide chains in the same manner as described for T5st(0). The results are shown in Table III. The T5-related phage *PB* appears to be equivalent to T5st(0). T5st(8) is just distinguishable from st(0).

T5 wild and st(14) show marked differences from st(0) and from each other.

These mutants were obtained from Irwin Rubenstein (Molecular Biology Department, Yale University) who has measured the relative DNA contents of these various mutants. Normalizing his relative numbers to our mean length of 38.8μ for T5st(0) one may calculate the total length of polynucleotide

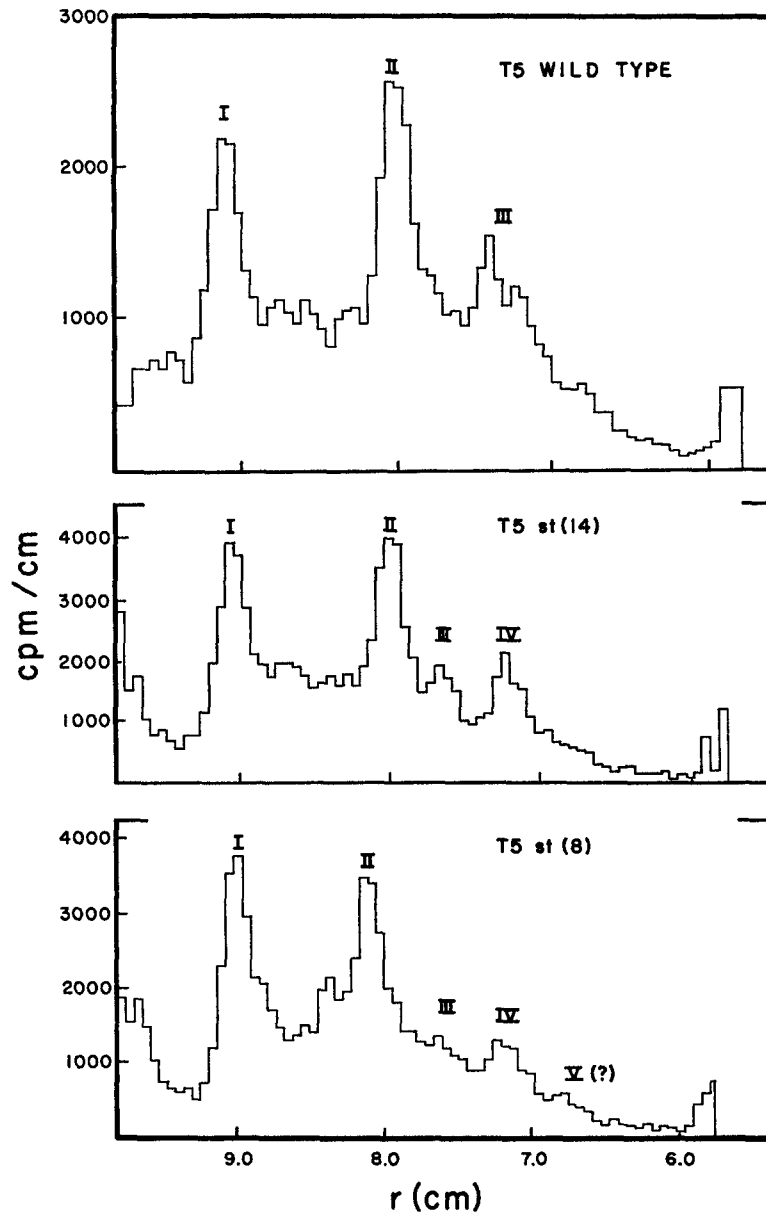


FIGURE 8. Comparison of alkaline sedimentation profiles for three T5 phage mutants.

chain in the three other cases. As shown in Table III, the total length of single chain determined by sedimentation agrees with the total length calculated from Rubenstein's data.

Thus the location of these gaps is under the control of the phage.

MODELS FOR T5ST(0) Turning back to the most studied T5st(0), there are a total of twelve possible models that will satisfy the data. The one that we favor is shown in Fig. 9. All of them contain a $4.0 + 34.0 \mu$ length in the first strand and various permutations of the remaining 14.4, 8, and 4μ lengths in the second strand.

TESTS OF MODEL This kind of model indicates that peak I contains most of *one* of the two complementary halves of the molecule while peak II contains

TABLE III
THE ANATOMY OF T5 MUTANT MOLECULES

Phage	No. of runs	I	II	III	IV	Sum of chain lengths	Sum from DNA content
T5 wild	4	36.9 (1)	12.1 (3)	5.4 (2)	Missing	84.8	83.2
T5st(14)	5	35.8 (1)	13.0 (2)	8.2 (1)	4.3 (3)	82.9	80.3
T5st(8)	2	34.2 (1)	14.6 (2)	9.5 (1)	3.6 (2)	80.1	77.8
T5st(0) and PB	9	34.0 (1)	14.4 (2)	8.0 (1)	4.0 (2)	78.8	77.6

Columns labelled I, II, III, and IV contain the length in microns and the number of chains of that length in parentheses; i.e., II—12.1 (3): 3 chains 12.1 μ long. The last column lists twice the lengths of the duplex T5 mutant on the basis of the DNA contents measured by Rubenstein (9). Rubenstein's relative values are standardized to a length of 38.8 μ for T5st(0). There appears to be a uniform 2 to 3% difference between these two numbers indicating some unresolved systematic error. In view of the origin of these numbers the agreement is unexpectedly good.

somewhat less of the *other* complementary half. This in turn suggests that the chains found in peaks I and II would anneal better with each other than they would by themselves. We have purified peaks I and II by resedimentation and tested this idea. Since we are dealing with very small amounts of DNA, it was necessary to perform tests which did not rely on optical melting. Two tests were finally made: chromatography on hydroxyapatite, which separates single chains from duplex ones (19, 20) and sensitivity to exonuclease I, which is highly specific for single chains (21). To summarize the results of work detailed elsewhere (18, 22), peaks I and II *do* appear to contain complementary chains while peaks I and II *individually* contain many fewer self-complementary sequences. Thus these tests support the models indicated in Fig. 9.

Perhaps the most direct way of testing the models shown in Fig. 9 is by visualizing with the electron microscope the duplex segments formed by a brief annealing treatment at low concentration (2 $\mu\text{g/ml}$). Under our protein

film conditions, single chains are seen very faintly or not at all. Therefore limited (bimolecular) annealing of separated chains should produce a peak of lengths at 4.0, 8.0, and 14.4 μ . The observed length distribution, seen in Fig. 10, shows at 4, 8, and 15 μ . These results stand in marked contrast to T2 and T7, which anneal to form a large fraction of full length molecules. The lengths of annealed T7 chains are shown on the right of Fig. 10. We don't consider this a rigorous experiment in itself in view of the sampling problems involved. However, it was a striking confirmation of this model and

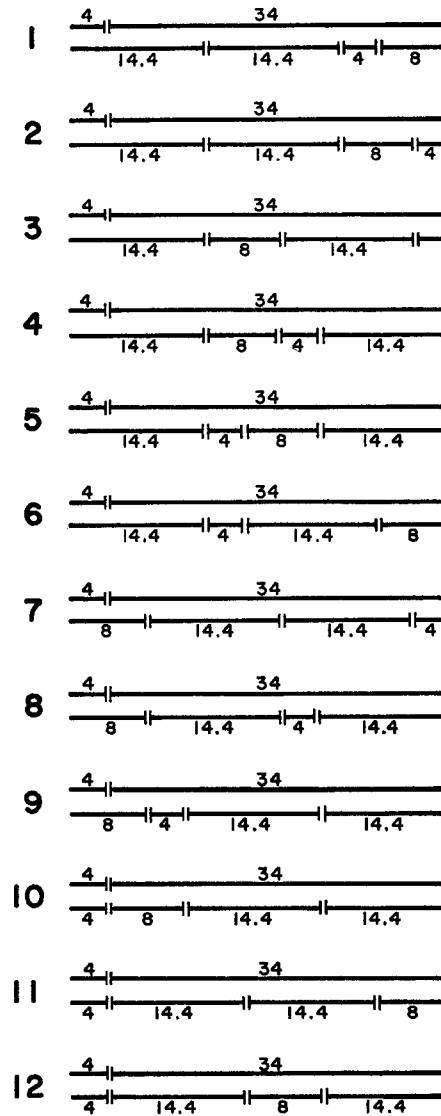


FIGURE 9. Models for the T5st(0) DNA molecule.

the work was completed long before the alkaline sedimentation profiles became available.

The T5 DNA molecule is the only known DNA molecule with unusual features along its length. What is the functional significance of these gaps in T5 DNA molecules? We have no answer to this question. Perhaps the 4.0 μ segment (at either end) corresponds to the first step transfer (FST) segment studied by Lanni (23). Perhaps the gaps (or the agents producing them) are responsible for the substantial unlinking of two groups of genetic markers in the T5 map (24).

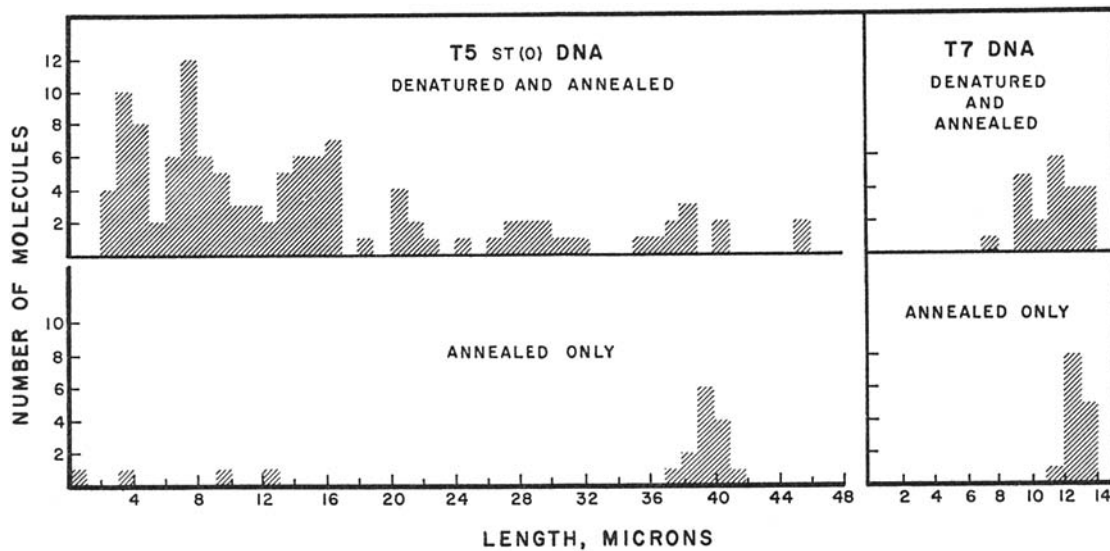


FIGURE 10. Distribution of molecular lengths seen after annealing T5 and T7 single chains.

We have done some preliminary experiments on the continuity of the T5 DNA molecules collected at various times after infection. To summarize briefly, the sedimentation rate of intracellular T5 DNA appears normal in neutral sucrose if the cells are broken open between 0 and 25 min after infection. However, by 4 min one can no longer find the characteristic four zone pattern in alkaline sucrose gradients: all the denatured DNA moves much faster. This makes us think that *all the gaps* are sealed very soon after infection. It will be interesting to learn exactly what agents are responsible for this closure and the subsequent reintroduction of specifically located gaps prior to or at maturation of the phage.

The Topography of DNA Molecules

There are two general ways of studying the topography of DNA molecules. One may perform genetic crosses, construct genetic maps, and then make

the assumption that this genetic map is a simple reflection of the physical structure of the DNA molecule. Alternatively, one may attempt to study the molecule directly. Of major importance for the interpretation of the genetic map is the proof that the genetic map and the genetic molecule are colinear and commensurate. To obtain this proof a direct study of the sequences in

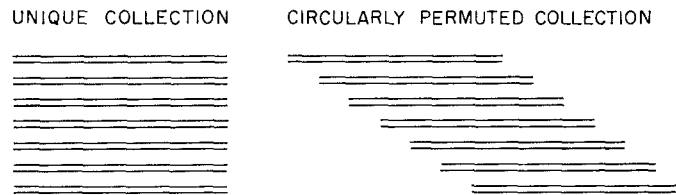


FIGURE 11. Unique and circularly permuted collection of DNA molecules.

DNA is required. This proof for λ dg was supplied in an elegant fashion by the work described by Dr. Hogness in this symposium.

Perhaps the most unusual and difficult situation is found in the most studied of all bacteriophage, the T-even. This phage engages in unusually vigorous recombination, it has a genetic map that is at least 500 map units long, and strikingly enough, it has a circular form (25). As mentioned before, T2 and T4 phage liberate *linear* DNA molecules 56 μ long. Thus the topology of the map and molecule disagree—to say nothing of topography. It was suggested by Stahl and Steinberg (26) that either the genetic recombination process was a peculiar two-event affair causing regions of the DNA molecule that were most distant from each other to be associated *or* that each phage particle contained molecules that were circular permutations of each other. The alternatives are shown in Fig. 11. We have been interested in devising tests to either prove or eliminate either possibility.

At the present time there are three different possible tests that one may apply to distinguish these two cases.

TEST 1 Break the DNA molecule into halves and quarters and test the fragments so obtained for some unique feature such as composition (density in CsCl) or transforming potential. This test has been applied to λ -DNA with success by Hogness and Simmons. The right and left halves obtained from a collection of λ -DNA molecules produce a bimodal band in CsCl gradients; they have different densities and composition. Therefore it is impossible for them to contain the same sequences (27). This fact is further secured by the demonstration that each half-molecule has a unique transforming ability (28). Thus there is little doubt that *most* of the λ -DNA molecules compose a unique collection of sequences. The experiments described by Dr. Hogness have demonstrated that the genetic map and λ -DNA mole-

cule are largely commensurate, thus vindicating this major genetic assumption.

With T5 and T2, no effective transformation has been available. When T2 is broken by shear, a bimodal band does not appear in the CsCl gradient. Even applying the most sensitive test that we could devise, Tom Pinkerton and I were unable to decide whether T2 were a unique or permuted collection (12). T5 DNA molecules defied analysis because half- and quarter-molecules had nearly the same density (29). Therefore a more general test was required.

TEST 2 Mechanically delete the middle region of the DNA molecules and test by annealing to see whether a special group of sequences has been removed. This test, devised with Irwin Rubenstein (30), is depicted in Fig. 12.

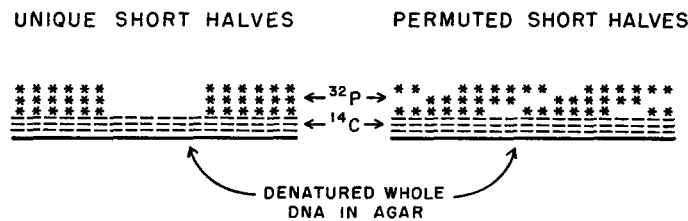


FIGURE 12. The agar column test for unique short halves. The heavy line denotes the denatured phage DNA molecule imbedded in agar. The dashes are ^{14}C -labeled single polynucleotide chain segments obtained by denaturing sonicated whole molecules. The stars are ^{32}P -labeled segments obtained by sonication and denaturation of "short halves"—or the smallest fragments collected from a MAK chromatogram of singly broken DNA molecules.

^{32}P -labeled DNA molecules are broken by shear which breaks rarely but most probably at the middle of the molecule. The majority of the molecules which are broken unequally are fractionated with respect to size on the MAK column. The shortest fragments are collected (the short halves). If the sequence of all molecules has the same starting point, then the short halves will be *missing* a certain fraction of their information. If the original collection were permuted, then all sequences would be found in the short halves. To test these alternatives, the ^{32}P -labeled short halves were mixed with ^{14}C -labeled whole molecules, sonicated to produce short segments, denatured, and annealed to denatured whole molecules imbedded in agar beads (31). As can be inferred from Fig. 12, under saturation conditions we would expect some fractionation of ^{14}C with respect to ^{32}P if the short halves were unique, while if they were permuted one would expect no fractionation because all segments would be equivalent.

The results of this test show clearly that T5 DNA molecules are a unique collection of sequences, while T2 DNA molecules are permuted. Therefore it is very likely that circular permutation among T2 DNA molecules is the basis of the circular genetic map.

TEST 3 Form artificial circular molecules by denaturation and annealing. Under these conditions intact single polynucleotide chains from a collection of permuted molecules will produce *circular* molecules which can be identified in the electron microscope. This test is depicted in Fig. 13. When

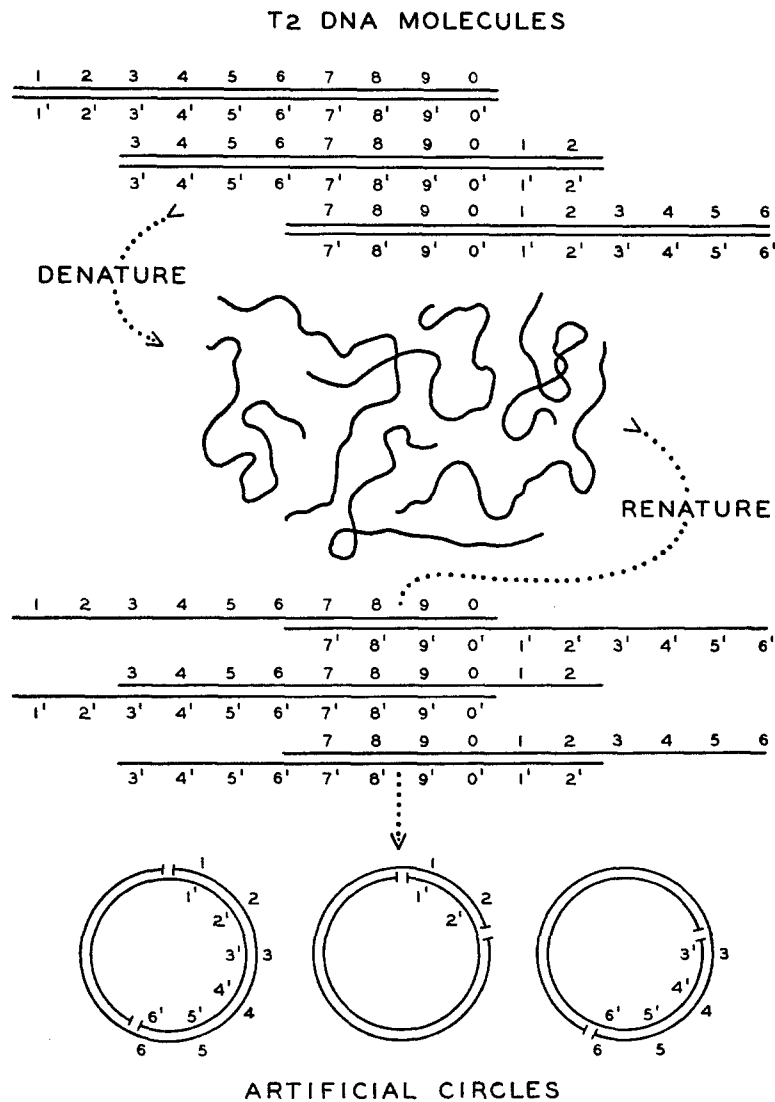


FIGURE 13. Circle formation from intact polynucleotide chains from circularly permuted T2(T4) DNA duplexes.

unbroken T2 or T4 DNA molecules ($1.25 \gamma/\text{ml}$) are treated with 0.20 M NaOH for 1 min (a treatment more than sufficient to separate the polynucleotide chains, 6), the pH returned to neutrality by dialysis and then annealed at 65°C for 40 min, a large fraction of the molecules are circular. An example of such a circular molecule is shown in Fig. 14. The linear molecules on the same grids show a wide variability in length that is expected, since only duplex regions are sufficiently visible with our procedure. These results are summar-

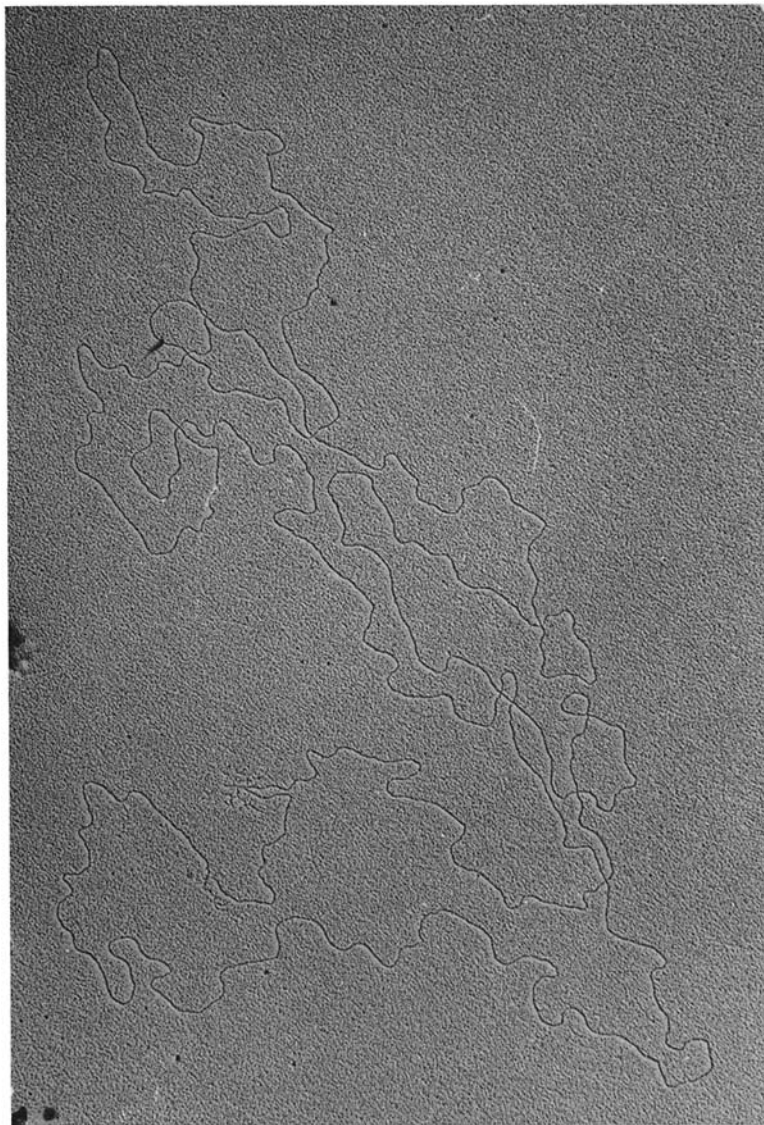


FIGURE 14. An artificial circle of T4 DNA. Notice the two "bushes" on the duplex molecule (lower right). These bushes appear at variable spacing on nearly all the most visible circles. They may be the terminally redundant single polynucleotide chains that can find no complementary chain. $\times 20,000$.

ized in Fig. 15. It is important to note that no circles have ever been found using half-molecules of T2 prepared by shear. This is in accord with expectation since the overlaps produced by chains of less than full length would not be complementary and the cooperation of four polynucleotide chains of the correct homology would be extremely unlikely.

Since the publication of this report (32) we have shown that both T2 and T4 can be cyclized in this manner either by alkali denaturation or by heating to 45°C in the presence of high concentrations of the denaturing anion ClO_4^-

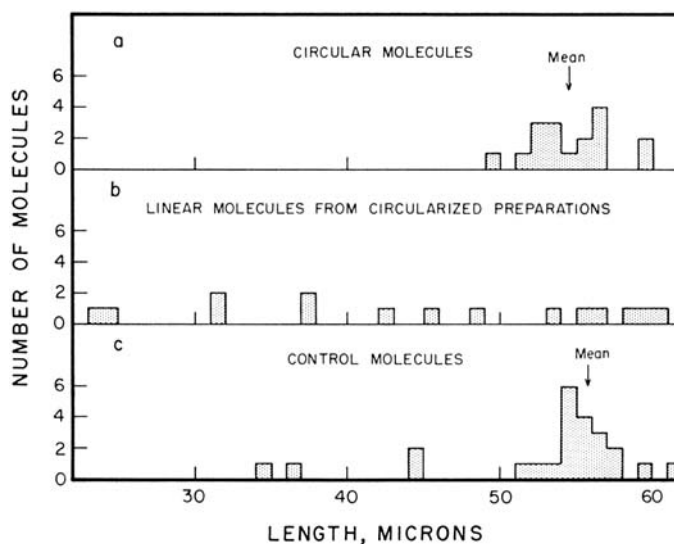


FIGURE 15. Histograms of measured contour lengths of duplex T2 DNA. (a) and (b), after denaturation and renaturation; (c) after renaturation treatment only.

(33). Annealing can be accomplished at 25°C in the presence of 7.2 M NaClO_4 —a fact which makes it possible to avoid the extended treatment at elevated temperatures which are known to cause chain cleavages (34). Apparently the annealing proceeds rapidly to form circles even after admixture of the basic protein, cytochrome *c*, indicating that protein conjugation with polynucleotide chains does not necessarily impede reformation of the duplex.

When this test is applied to other phage DNA's, the following facts are clear. T5 DNA will not produce circles by this treatment: this is expected since the collection is unique; moreover, the four specific gaps would make it improbable that circles could be formed even if the collection were permuted. T7 DNA has been tested and no circles have been found; therefore, we think the collection is unique. P22, the temperate bacteriophage used in transductional analysis of *Salmonella*, will cyclize upon annealing after being exposed to denaturing alkaline pH, but not before. On the basis of this experiment we think that P22 liberates a circularly permuted collection of

DNA molecules and should show a circular genetic map (38). If this proves to be true, then the T-evens must no longer be considered as special cases. DNA molecules from other phages are currently being studied.

DNA Molecules with the Same Sequence at Both Ends

On the basis of some genetic arguments it was proposed that T2-T4 DNA molecules contain a short region at the end of each molecule, the sequence of which is the same as the sequence at the beginning (25). This terminal redundancy was hypothesized to account for heterozygotes of deletion mutations which might be expected to encounter mechanical difficulties if present as an internal heteroduplex (35-37). We decided to attempt to obtain direct physical evidence for this terminal redundancy by partially digesting the molecule with exonuclease III, the *coli* nuclease that removes nucleotides stepwise from the 3'-ends of duplex molecules. The resulting molecule has 5'-ended single chains left at each end. If the molecule were terminally redundant, then these exposed polynucleotide chains would be mutually complementary and would be expected to anneal to form circles. The experimental plan and the expected results are shown in Fig. 16. In collaboration with

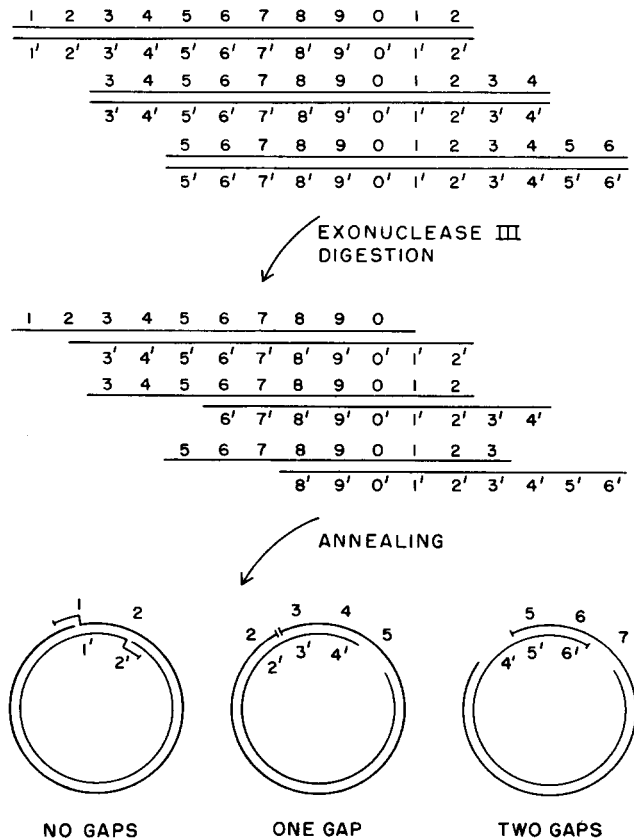


FIGURE 16. Experimental plan to test for terminal redundancy.

Dr. Richardson of Harvard and Drs. Ritchie and MacHattie in our laboratory, this experiment was done and we have found that digestion with exonuclease III renders T*2 molecules cyclizable. One example of more than 50 circles is shown in Fig. 17. Our present inventory is summarized in Table IV.

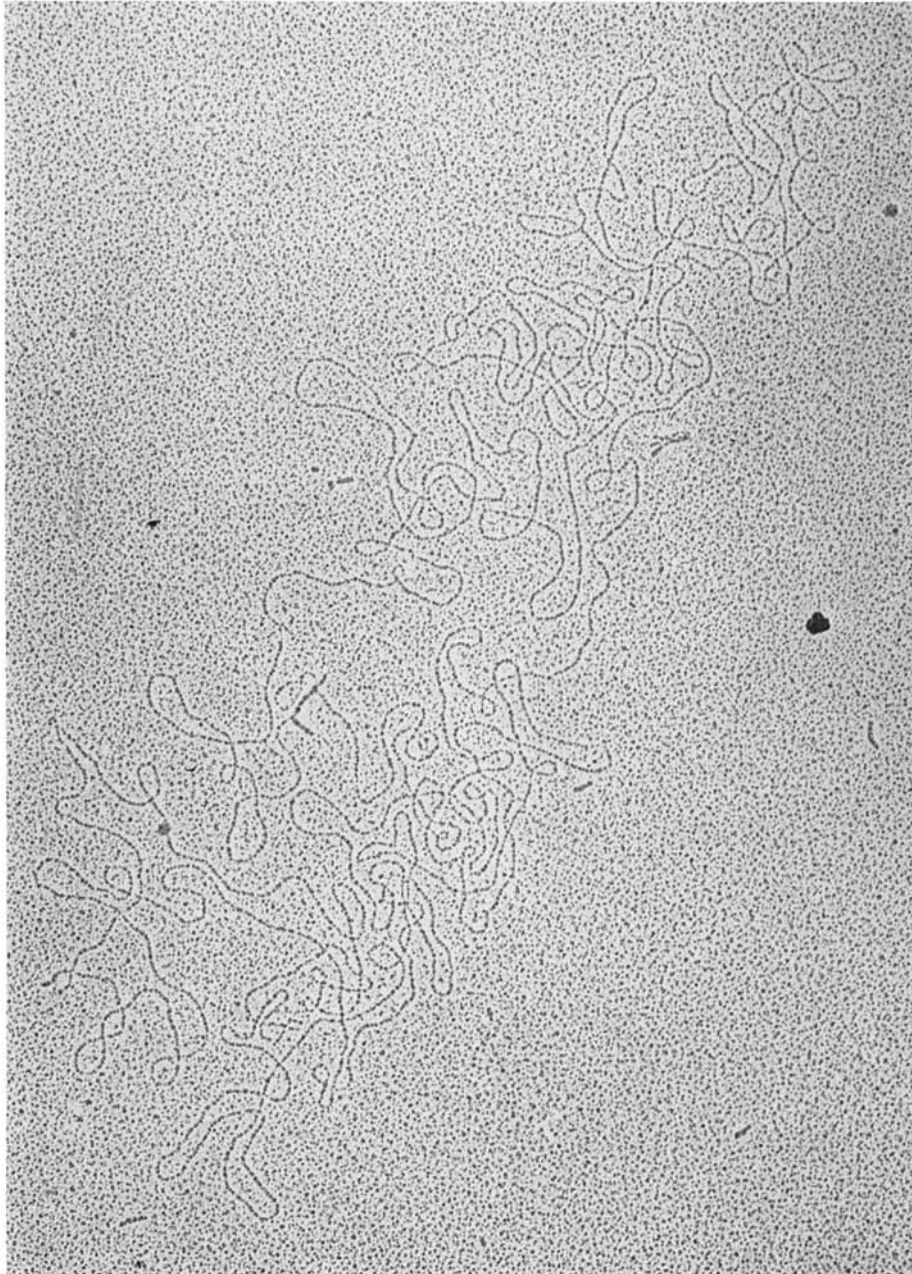


FIGURE 17. An artificial circle produced by annealing a T*2 DNA molecule partially digested with exonuclease III. $\times 44,000$.

TABLE IV
CIRCLE FORMATION AS RELATED TO PER CENT DIGESTION

Digestion	Linear	Circles	Uncertain	Total	Circles
<i>per cent</i>					<i>per cent</i>
0	95	0	5	100	0
0.92	91	1	3	95	1.1
3.6	119	32	24	175	18.3
4.7	54	9	3	66	13.6
8.0	22	1	0	23	4.5

As seen in this table and more clearly in Fig. 18, the frequency of circles is *zero* at 0% digestion, reaches a maximum at about 3%, and falls rapidly as more digestion takes place. We think that the reason for this decline is that circles are formed, but that they are not recognized as such.

GAPS As shown in the experimental plan (Fig. 16), if the digestion proceeds beyond the length of the terminal redundancy, then the circle will contain a single-chain region for which a complement is unavailable. Since we do not see single chains very well, this results in a circle with a "gap" in it. A circle with a gap in it is logically a linear molecule, which on technical grounds should not be included among our collection of circular molecules. We have compromised in this way: some of the molecules (13/48) have no gaps, therefore they present no problem. A gapped molecule is identified as a circle if: (a) the gap is short and the ends point toward each other; (b) if a faint "single chain" can be seen connecting two ends. Of particular interest are the molecules containing two gaps. The duplex segment between the two gaps should correspond to the length of the terminally redundant region. At

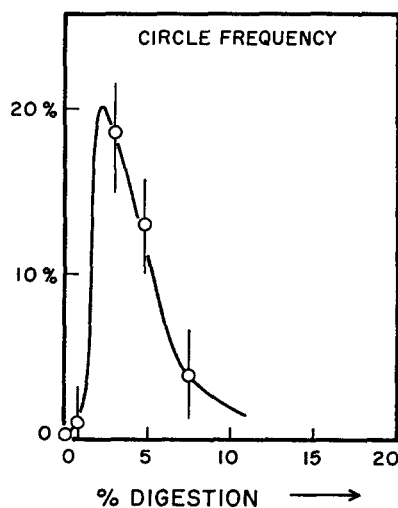


FIGURE 18. The frequency of circles as a function of % of nucleotides removed by exonuclease III. The fraction of nucleotides rendered acid-soluble by exonuclease III was determined and aliquots were removed at various times and annealed in 7.2 M NaClO₄ at 25°C, and examined in the protein film.

present, we have two such molecules with intergap distances 1.7 and 0.6 μ , which give 3.2% and 1.1% respectively when divided by $53 \pm 2 \mu$, the contour length of the linear or circular molecule. Thus our preliminary estimate of the average length of the terminal redundancy is 1 to 3%. The results appear to be in accord with the genetic estimates of the terminal redundancy, which are about 1 to 2%.

Of crucial importance in this experiment is the demonstration that half or smaller length molecules do *not* form circles after exonuclease III digestion. At the present time, we have never seen a convincing circle the contour length of which is not equivalent to an *intact* molecule. This is true even though our preparations did contain fragments. We have further confidence that this control will hold up since T2 half-molecules which were melted and annealed as part of test 3 for circular permutation did not produce circles that could be found. These overlapped molecules would have much the same structure as exonuclease III-digested half-molecules.

Thus T2 DNA molecules are a circularly permuted collection of linear duplex sequences, many of which contain terminally redundant sequences. This is in exact accord with the hypothesis of Stahl and Steinberg (26). This is a welcome solution to a long standing problem. There is every reason to believe that the genetic map and the DNA molecule are colinear and commensurate; indeed, the unusual nature of T2 DNA supports this simple and widely held view.

CONCLUSIONS

1. Most phage contain linear duplex DNA molecules, the length of which can be reliably measured in the electron microscope.
2. Observed molecular length correlates with molecular weight if the molecules have a linear density of 196 daltons/A—that of the B form of DNA.
3. With the exception of T5 and PB bacteriophage, all the other phages studied (T2, T*2, T4, P1, P22, λ , and T7) contain polynucleotide chains that are continuous over most of the length of the molecule.
4. T5st(0) DNA molecules contain four gaps: one in the chain A and three in the chain B. They appear to be true interruptions.
5. The location of these gaps is under the genetic control of the phage, not of the host cell, because different phage mutants produce single chains of different lengths.
6. By means of two completely different kinds of tests, it was found that T2(T4) DNA molecules consist of a circularly permuted collection of sequences.
7. Partial (3%) digestion of T*2 DNA with exonuclease III predisposes a large fraction of the molecules to form circles upon annealing. This is in exact accord with the hypothesis of terminal redundancy.

8. We visually estimate the extent of this redundancy to be about 1 to 3% of the molecule, which is in accord with genetic estimates.

It is a pleasure to acknowledge my collaborators, for whom I am the spokesman. Dr. Lorne MacHattie has made much of the electron microscope work possible. The anatomy of T5 DNA was largely dissected by Dr. John Abelson. Mr. Marcus Rhoades has made a number of decisive contributions to this and other problems. Dr. Don Ritchie together with Dr. MacHattie has had the major responsibility for the terminal redundancy work. We are also indebted to Dr. Irwin Rubenstein who has given us his current information regarding T5 and T7 DNA molecules. Finally, we thank our benefactors: the Atomic Energy Commission (AT(30-1)2119), The National Institutes of Health (E-3233), and The National Science Foundation (G-10726) and the general support of the Biophysics Department at The Johns Hopkins University.

REFERENCES

1. KLEINSCHMIDT, A., LANG, D., and ZAHN, R. K., Über die intrazelluläre Formation von Bakterien-DNS, *Z. Naturforsch.*, 1961, **16**, 730.
2. WILKINS, M. H. F., Molecular configuration of nucleic acids, *Science*, 1963, **140**, 3570.
3. LEVINTHAL, C., and THOMAS, C. A., Molecular autoradiography: The β -ray counting from single virus particles and DNA molecules in nuclear emulsions, *Biochim. et Biophysica Acta*, 1957, **51**, 32.
4. RUBENSTEIN, I., THOMAS, C. A., and HERSHEY, A. D., The molecular weights of T2 bacteriophage DNA and its first and second breakage products, *Proc. Nat. Acad. Sci.* 1961, **47**, 1113.
5. BURGI, E., and HERSHEY, A. D., Sedimentation rate as a measure of molecular weight of DNA, *Biophysic. J.*, 1963, **3**, 309.
6. STUDIER, F. W., Sedimentation studies of the size and shape of DNA, *J. Mol. Biol.*, 1965, **11**, 373.
7. MACHATTIE, L. A., and THOMAS, C. A., DNA from bacteriophage lambda: molecular length and conformation, *Science*, 1964, **144**, 1142.
8. KLEINSCHMIDT, A., BURTON, A., and SINSHEIMER, R. L., Electron microscopy of the replicative form of the DNA of the bacteriophage ϕ x174, *Science*, 1963, **142**, 961.
9. RUBENSTEIN, I., and LEIGHTON, S. B., personal communication.
10. ENOMOTO, S., and KRIMM, S., Elastic moduli of helical polypeptide chain structures, *Biophysic. J.*, 1962, **2**, 317.
11. BERNS, K. I., and THOMAS, C. A., A study of single polynucleotide chains derived from T2 and T4 bacteriophage, *J. Mol. Biol.*, 1961, **3**, 289.
12. THOMAS, C. A., and PINKERTON, T. C., Sedimentation equilibrium studies on intact and fragmented bacteriophage DNA, *J. Mol. Biol.*, 1962, **5**, 356.
13. THOMAS, C. A., PINKERTON, T. C., and RUBENSTEIN, I., The structure and intramolecular density heterogeneity of the T2 phage DNA molecule, in *Informational Macromolecules*, (V. Bryson and H. Vogel, editors), New York, Academic Press Inc., 1963, 89.
14. VINOGRAD, M. J., DAVIDSON, N., and DOVE, W. F., JR., The buoyant behavior of viral and bacterial DNA in alkaline CsCl, *Proc. Nat. Acad. Sci.*, 1963, **49**, 12.

15. BLATTNER, F., ABELSON, J., PINKERTON, T. C., and THOMAS, C. A., JR., in preparation.
16. RHOADES, M., ABELSON, J., and THOMAS, C. A., The question of interruptions in bacteriophage DNA molecules, *Abstr. 9th Ann. Biophysics Soc. Meetings*, 1965, 96.
17. TOMIZAWA, J., and ANRAKU, N., Molecular mechanisms of genetic recombination in bacteriophage. IV. Absence of polynucleotide interruption in DNA of T4 and λ phage particles, with special reference to heterozygosis, *J. Mol. Biol.*, 1965, **11**, 509.
18. ABELSON, J., The Anatomy of the T5 DNA Molecule, Doctoral Thesis, The Johns Hopkins University, 1965.
19. BERNARDI, G., Chromatography of nucleic acids on hydroxyapatite, *Nature*, 1965, **206**, 779.
20. MIYAZAWA, Y., and THOMAS, C. A., Nucleotide composition of short segments of DNA molecules, *J. Mol. Biol.*, 1965, **11**, 223.
21. LEHMAN, I. R., The nucleases of *Escherichia coli*, in *Progress in Nucleic Acid Research*, (J. N. Davidson and W. E. Cohn, editors), New York, Academic Press Inc., 1963, 84.
22. ABELSON, J., and THOMAS, C. A., The anatomy of the T5 bacteriophage DNA molecule, *J. Mol. Biol.*, 1966, in press.
23. LANNI, Y. T., MCCORQUODALE, D. J., and WILSON, C. M., Molecular aspects of DNA transfer from phage T5 to host cells. II. Origin of first-step-transfer DNA fragments, *J. Mol. Biol.*, 1964, **10**, 19.
24. FATTIG, W. D., and LANNI, F., Mapping of temperature-sensitive mutants in bacteriophage T5, *Genetics*, 1965, **51**, 157.
25. STREISINGER, G., EDGAR, R. S., and DENHARDT, G. H., Chromosome structure in phage T4. I. Circularity of the linkage map, *Proc. Nat. Acad. Sc.*, 1964, **51**, 775.
26. STAHL, F., and STEINBERG, C., The theory of formal phage genetics for circular maps, *Genetics*, 1964, **50**, 531.
27. HOGNESS, D. S., and SIMMONS, J. R., Breakage of λ dg DNA: Chemical and genetic characterization of each isolated half-molecule, *J. Mol. Biol.*, 1964, **9**, 411.
28. KAISER, A. D., The production of phage chromosome fragments and their capacity for genetic transfer, *J. Mol. Biol.*, 1962, **4**, 275.
29. THOMAS, C. A., and PINKERTON, T. C., unpublished experiments.
30. THOMAS, C. A., and RUBENSTEIN, I., The arrangements of nucleotide sequences in T2 and T5 bacteriophage DNA molecules, *Biophysic. J.*, 1964, **4**, 93.
31. HOYER, B. H., MCCARTHY, B. J., and BOLTON, E. T., A molecular approach in the systematics of higher organisms, *Science*, 1964, **144**, 959.
32. THOMAS, C. A., and MCHATTIE, L. A., Circular T2 DNA molecules, *Proc. Nat. Acad. Sc.*, 1964, **52**, 1297.
33. GEIDUSCHEK, E. P., On the factors controlling the reversibility of DNA denaturation, *J. Mol. Biol.*, 1962, **4**, 467.
34. EIGNER, J., and DOTY, P., The native, denatured and renatured states of deoxyribonucleic acid, *J. Mol. Biol.*, 1965, **12**, 549.

35. LEVINTHAL, C., Recombination in phage T2: Its relationship to heterozygosis and growth, *Genetics*, 1954, **39**, 169.
36. SÉCHAUD, J., STREISINGER, G., EMRICH, J., NEWTON, J., LANFORD, H., REINHOLD, H., and STAHL, M. M., Chromosome structure in phage T4. II. Terminal redundancy and heterozygosis, *Proc. Nat. Acad. Sci.*, 1965, **54**, 1333.
37. SHALITIN, C., and STAHL, F., Additional evidence for two kinds of heterozygotes in T4, *Proc. Nat. Acad. Sc.*, 1965, **54**, 1340.
38. LEVINE, M., and CURTISS, R., Genetic fine structure of the C region and the linkage map of phage P22, *Genetics*, 1961, **46**, 1573.