9 August 1968, Volume 161, Number 3841 SCIENCE

Repeated Sequences in DNA

Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms.

R. J. Britten and D. E. Kohne

The complementary structure of DNA plays a fundamental role in the cell. The complementary relations between nucleotide pairs are important not only in the duplication of DNA, but in the transcription and translation of genetic information. Matching of complementary nucleotide sequences is probably involved in genetic recombination as well as in other events of recognition and control within the cell.

It is a remarkable fact that separated complementary strands of *purified* DNA recognize each other. Under appropriate conditions they specifically reassociate (*1*). This phenomenon has supplied a useful tool for exploring the nature of molecular events within the cell and broader biological questions such as the relationships among species (*1–3*).

Simple complementary ribopolymer pairs were shown in 1957 to form a helical paired structure when mixed in solution (*4*). In 1960, DNA was dissociated into two strands, and the physical properties and biological activity of double-stranded DNA were then restored by incubation under appropriate conditions (*1–5*). In 1961, virus-specific RNA, made by bacteria during viral infection, was shown to pair with the viral DNA (*6*). Techniques were developed for the immobilization of single-stranded DNA in cellulose (*7*), in agar (*8*), and on nitrocellulose filters (*9*). It then became possible to assay the reassociation of radioactively labeled single-stranded fragments of DNA or RNA with the immobilized DNA.

Reassociation of the DNA of vertebrates was observed in 1964 (*3*). The extent of reassociation between DNA strands derived from different species was shown to be a measure of the evolutionary relation between the species (*10*). However, measurements also showed that the nucleotide sequence pairing was imprecise even when DNA from a single species was reassociated (*11*).

Before these measurements were made it had been expected that it would be very difficult to observe the reassociation of the DNA of vertebrates and other higher organisms (*1*). The enormous dilution of individual nucleotide sequences in the large quantity of DNA in each cell was expected to make the reaction so slow that months would be required for its completion at practical concentrations with the DNA-agar method.

Investigation of this paradox was begun in our laboratory in early 1964, and shortly afterward the hypothesis was put forward (*12*) that some nucleotide sequences were frequently repeated in the DNA of vertebrates. This supposition was supported by the observation that 10 percent of the DNA of the mouse reassociated extremely rapidly. This fraction, identified as mouse-satellite DNA was shown by later measurements (*13*) to consist of a million copies of a short nucleotide sequence (*13*). Later work (*14, 15*) has shown that repeated nucleotide sequences are of very general occurrence.

In this article we describe selected measurements (*12–15*) that show most clearly the presence of repeated sequences and indicate some of their properties.

Conditions for Reassociation

The conditions for efficient reassociation of DNA were explored originally by Marmur *et al.* (*1*) and have since been studied in several laboratories (*15–17*). Briefly stated, the requirements are as follows. (i) There must be an adequate concentration of cations. Below 0.01*M* sodium ion, the reassociation reaction is effectively blocked. (ii) The temperature of incubation must be high enough to weaken intrastrand secondary structure. The optimum temperature for reassociation is about 25°C below the temperature required for dissociation of the resulting double strands. (iii) The incubation time and the DNA concentration must be sufficient to permit an adequate number of collisions so that the DNA can reassociate. (iv) The size of the DNA fragments also affects the rate of reassociation and is conveniently controlled if the DNA is "sheared" to small fragments (*18*). Thus, in order to achieve reproducible reassociation reactions the cation concentration, temperature of incubation, DNA concentration, and DNA fragment size must all be controlled (*19*).

The Measurement of Reassociation

Reassociation can be measured in a variety of ways, each depending on some easily detected physical difference between single-stranded (dissociated) DNA and double-stranded (reassociated) DNA (*1*). For example, dissociated DNA absorbs more ultraviolet light than reassociated DNA does. Double-strand DNA also has a greater

The authors are staff members of the Department of Terrestrial Magnetism of the Carnegie Institution of Washington, Washington, D.C. 20015.

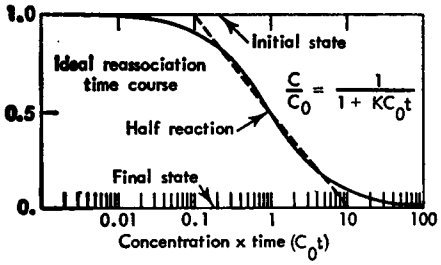


Fig. 1. Time course of an ideal, second-order reaction to illustrate the features of the log C_0t plot. The equation represents the fraction of DNA which remains single-stranded at any time after the initiation of the reaction. For this example, K is taken to be 1.0, and the fraction remaining single-stranded is plotted against the product of total concentration and time on a logarithmic scale.

degree of optical activity than single-strand DNA.

In the DNA-agar method, reassociation is monitored by measuring the binding of labeled fragments of single-stranded DNA to long strands of DNA physically immobilized in a supporting substance. The immobilization prevents reassociation of the long DNA with itself. After incubation the unbound fragments are washed away, and the quantity of bound radioactive fragments is measured. It is now possible to measure the reassociation of DNA fragments with DNA immobilized on nitrocellulose filters (*20*). The rate of the reaction is markedly reduced compared to the rate in solution (*21*).

Another useful technique for measuring reassociation depends on the fact that double-stranded DNA can be separated from single-stranded DNA on a calcium phosphate (hydroxyapatite) column (*22*). Reassociation reactions can be followed by passing samples through hydroxyapatite and determining the amount of double-stranded DNA adhering to the column. This technique is particularly useful since DNA can be fractionated on a preparative scale on the basis of its ability to reassociate at a given C_0t, a parameter which may be explained as follows.

The Meaning of C_0t

Much of the evidence for repeated sequences depends on measurements of the rate of reassociation. In addition, the design of most reassociation experiments is strongly influenced by the time required to complete the process. The reassociation of a pair of complementary sequences results from their collision, and therefore the rate depends on their concentration. The product of the DNA concentration and the time of incubation is the controlling parameter for estimating the completion of a reaction. For convenience and simplification of language we have chosen to call this useful parameter C_0t, which is expressed in moles of nucleotides times seconds per liter (*23*).

Evidence is presented below that the DNA of each organism may be characterized by the value of C_0t at which the reassociation reaction is half completed under controlled conditions. The rates observed range over at least eight orders of magnitude. Therefore we have found it necessary to introduce a simple logarithmic method for the presentation of measurements of reactions over extended periods of time and wide ranges of concentration. For illustration, Fig. 1 shows the progress of an ideal second-order reaction plotted as a function of the product of the time of reaction and the DNA concentration on a logarithmic scale. On such a graph, reactions carried out at different concentrations may be compared, and the data may be combined to give a complete view of the time course of the reaction.

The symmetrical shape of an ideal second-order curve plotted in this way is pleasing and convenient. The central two-thirds of the curve follows closely a straight line, shown dashed. One useful indicator is the slope of this line which can be evaluated from the ratio of the values of C_0t at its two ends. This ratio is about 100 for an ideal reaction when estimated as shown on Fig. 1. If the ratio is much greater than 100, the reaction is surely heterogeneous; that is, species with widely different rates of reassociation are present.

Rate of Reassociation of DNA

Reassociation of a pair of complementary strands results from their collision. Therefore we expect the half-period for reassociation to be inversely proportional to the DNA concentration under fixed conditions for a particular DNA (*24*).

Further, one would expect for a given total DNA concentration that the half-period for reassociation would be proportional to the number of different types of fragments present and thus to the genome size (*25*). This expectation is exactly borne out in several cases. In Fig. 2, the time course of reassociation of a number of double-stranded nucleic acids is shown. Within the precision of

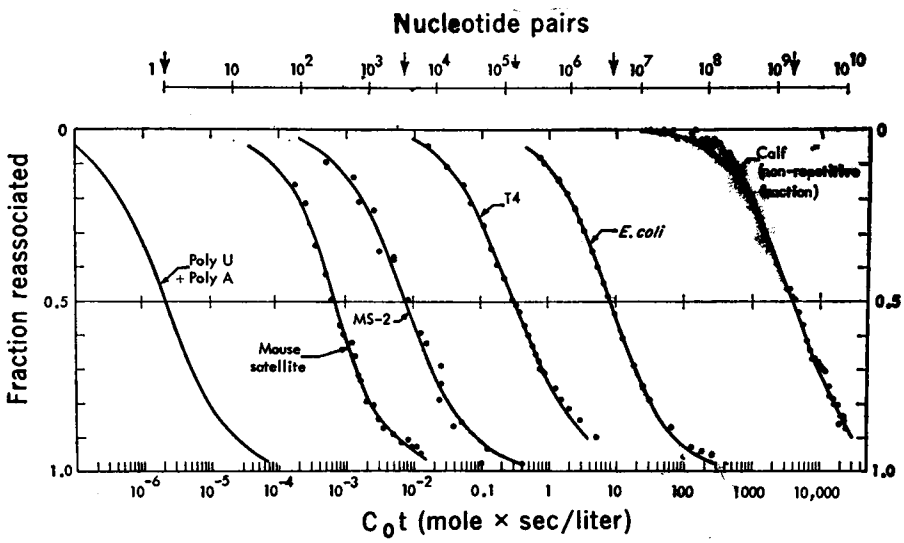
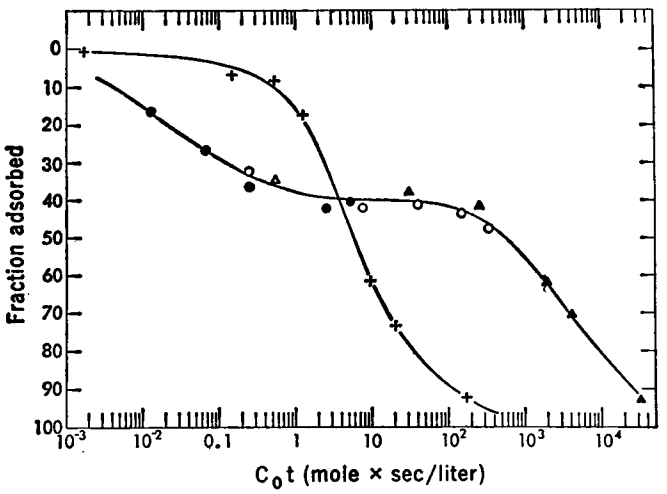


Fig. 2. Reassociation of double-stranded nucleic acids from various sources. The genome size (*25*) is indicated by the arrows near the upper nomographic scale. Over a factor of 10^9, this value is proportional to the C_0t required for half reaction. The DNA was sheared (*18*) and the other nucleic acids are reported to have approximately the same fragment size (about 400 nucleotides, single-stranded). Correction has been made (*19*) to give the rate that would be observed at 0.18*M* sodium-ion concentration. No correction for temperature has been applied as it was approximately optimum in all cases. Optical rotation was the measure of the reassociation of the calf thymus nonrepeated fraction (far right). The MS-2 RNA points were calculated from a series of measurements (*28*) of the increase in ribonuclease resistance. The curve (far left) for polyuridylic acid + polyadenylic acid was estimated from the data of Ross and Sturtevant (*29*). The remainder of the curves were measured by hypochromicity at 260 nm; a Zeiss spectrophotometer with a continuous recording attachment was used.



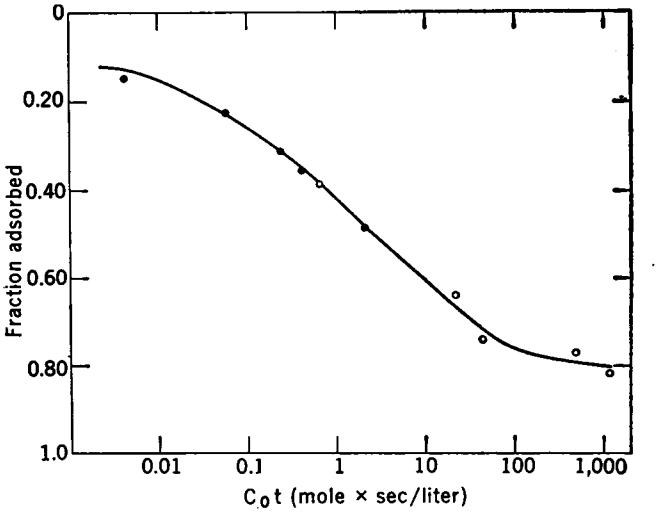


Fig. 3 (left). The kinetics of reassociation of calf-thymus DNA measured with hydroxyapatite. The DNA was sheared at 3.4 kilobars (*18*) and incubated at 60°C in 0.12*M* phosphate buffer (*31*). At various times, samples were diluted, if necessary (in 0.12*M* phosphate buffer at 60°C), and passed over a hydroxyapatite column at 60°C. The DNA concentrations during the reaction were (μg/ml): open triangles, 2; closed circles, 10; open circles, 600; closed triangles, 8600. Crosses are radioactively labeled *E. coli* DNA at 43 μg/ml present in the reaction containing calf thymus DNA at 8600 μg/ml. Fig. 4 (right). The kinetics of reassociation of salmon sperm DNA measured with hydroxyapatite. The DNA was sheared at 3.4 kilobars and incubated at 50°C in 0.14*M* phosphate buffer. The samples were diluted into 0.14*M* phosphate buffer at 50°C and passed over hydroxyapatite at 50°C. The DNA concentrations during the incubation were (μg/ml): closed circles, 8; open circles, 1600.

the measurements, the reassociation of these various DNA's follows the time course of a single second-order reaction. In each case where it is applicable, the genome size (*25*) is marked with an arrow on the upper scale. Cairn's measurement (*26*) of the size of the *Escherichia coli* genome (4.5×10^6 nucleotide pairs) has been used to fix and locate this scale. The length of the T2 bacteriophage chromosome has also been carefully measured and found to be 2×10^5 nucleotide pairs, and the size of T4 is similar (*27*). The total length of MS-2 viral RNA is 2.4×10^6 or 4000 nucleotide pairs in the double-stranded replicative form (*28*). The rate of reassociation (*29*) of the homopolymer pair [polyuridylic acid plus polyadenylic acid (polyU + polyA)] is consistent with the fact that these molecules are complementary in all possible registrations.

The proportionality between the C_0t (*23*) required for half-reassociation of the DNA and the genome size (*25*) is only true in the absence of repeated DNA sequences.

Figure 2 also shows the time course of reassociation for two fractions isolated from mammalian DNA. These fractions both follow the curve expected for a single second-order reaction, but one fraction reassociates more rapidly than the smallest virus, while the other reassociates 500 times more slowly than bacterial DNA. The former (mouse satellite DNA) represents 10 percent of the mouse DNA; its rate of reassociation indicates that the segment is roughly 300 nucleotide pairs and must be repeated about a million times (*13*) in a single cell. At the other extreme is a slowly reassociating fraction which includes about 60 percent of calf DNA. Its rate of reassociation is just that expected if it were made of unique (non-repeating) sequences. The calf genome contains 3.2×10^9 nucleotide pairs (*30*).

Repeated Sequences in the DNA of Calf and Salmon

In order to obtain a fairly complete view of the repeated sequences in one organism, it is necessary to measure the degree of reassociation over a very wide range of C_0t. In Fig. 3 the reassociation of calf thymus DNA measured by the hydroxyapatite procedure is shown. The hydroxyapatite method is convenient for this purpose since the degree of reassociation can be directly determined by assay of the amount of DNA which is bound (*22, 15*). Samples were simply diluted into a convenient volume of 0.12*M* phosphate buffer (*31*) and passed over hydroxyapatite in a water-jacketed column at 60°C. A variety of tests (*15, 32*) have shown that under these conditions reassociated DNA is quantitatively bound, while not more than ½ percent of single-stranded DNA is adsorbed. The concentration of DNA present in the incubation mixture also can be varied over a wide range without interfering with the determination.

The hydroxyapatite binding measurements (Fig. 3) show that 40 percent of the calf DNA has reassociated before a C_0t of 2. Little if any reaction occurs in the next two decades of C_0t. Thus for calf DNA, there is a clear separation between DNA which reassociates very rapidly and that which reassociates very slowly.

The rapidly reassociating fraction in calf DNA requires a C_0t of 0.03 for half-reassociation, whereas the slowly reassociating fraction requires a C_0t of 3000. Thus the concentration of DNA sequences which reassociate rapidly is 100,000 times the concentration of those sequences which reassociate slowly. If the slow fraction is made up of unique sequences, each of which occurs only once in the calf genome, then the sequences of the rapid fraction must be repeated 100,000 times on the average.

The measurements shown on Fig. 3 were done in several series at different DNA concentrations. Nevertheless, the results are concordant. The points fall on a single curve with good accuracy. This establishes that the measured reassociation process results from a bimolecular collision. In turn, the rapidity of the early part of the reassociation reaction can result only from high concentrations of the reacting species. We

may conclude that about 40 percent of calf DNA consists of sequences which are repeated between 10,000 and a million times.

In one of the series of measurements shown on Fig. 3, in addition to the 8600 micrograms of sheared calf DNA per milliliter there was present 43 micrograms of P^{32}-labeled sheared *E. coli* DNA per milliliter, serving as an "internal standard." The simultaneous assay of the reassociation of the two DNA's that are present together controls a variety of possible experimental errors (*33*).

Figure 4 shows the reassociation of salmon sperm DNA measured with hydroxyapatite. Most of the salmon DNA appears to be made up of repeated sequences. The average degree of repetition is not as great as it is for the repeated fraction of calf DNA. The fact that the major part of the process extends over more than a factor of 10,000 in C_0t shows that many different degrees of repetition are present, varying from perhaps 100 copies to as many as 100,000 copies. The reassociation of the unique (single copy) DNA of this organism has not yet been observed. It would be expected to reassociate with a C_0t greater than 1000, and no measurements have yet been made in this region for salmon DNA.

The Occurrence of Repetitious DNA

With the observation of repeated DNA sequences in several vertebrate genomes the question arises: Are the DNA's of these creatures exceptional, or do repeated sequences occur generally among higher organisms? A limited survey was therefore carried out by the following procedure. (i) DNA was prepared and purified (*34*) from many organisms and then sheared to fragments consisting of 500 nucleotides (*18*). (ii) The DNA was dissociated in 0.12*M* phosphate buffer (*31*) and incubated at C_0t of 1 to 10 at 60°C. (iii) The solution was passed over a hydroxyapatite column equilibrated to 0.12*M* phosphate buffer and 60°C (*15, 32*). Under these conditions only the reassociated DNA becomes adsorbed to the column. (iv) The adsorbed DNA was eluted, and its reassociation kinetics were measured in the spectrophotometer.

The optical reassociation measurements for rapidly reassociating fractions prepared in this way from DNA of four organisms are shown in Fig. 5. All of these organisms contain repetitive DNA. The rate of reassociation of these fractions is very much faster than that calculated from the respective genome sizes. However, the reassociation pattern observed is quite different in each of the four cases. The curves for sea urchin DNA and calf DNA are probably representative of most of the repeated DNA in these organisms. However, the curve for mouse satellite DNA represents only the most repetitive fraction of the DNA from mouse cells (*13*), since the C_0t used before fractionation was very small. In the case of the onion DNA there is another repeating fraction which reacts more slowly than the fraction from onion shown in Fig. 5. This more slowly reacting fraction appears to have a repetition frequency between 100 and 1000.

Table 1. Occurrence of repititious DNA.

- Protozoa
  - Dinoflagellate (*Gyrodinium cohnii*)*
  - *Euglena gracilis**
- Porifera
  - Sponge (*Microciona*)*
- Coelenterates
  - Sea anemone (*Metridium*) (tentacles)*
- Echinoderms
  - Sea urchin (*Strongylocentrotus*) (sperm)*†‡
  - Sea urchin (*Arbacia*) (sperm)*†‡
  - Starfish (*Asterias*) (gonads)*
  - Sand dollar (*Echinarachnis*)‡
- Arthropods
  - Crab (*Cancer borealis*) (gonads)*
  - Horseshoe crab (*Limulus*) (hepatopancreas)*
- Mollusks
  - Squid (*Loligo pealii*) (sperm)*
- Elasmobranchs
  - Dogfish shark (liver)*
- Osteichthyes
  - Salmon (sperm)*†‡
  - Lungfish*†
- Amphibians
  - Amphiuma (liver, red blood cells, muscle)*
  - Frog (*Rana pipiens*)†
  - Frog (*Rana sylvatica*)‡
  - Toad (*Xenopus laevis*) (heart, liver, red blood cells)
  - Axolotl (*Ambystoma tigrinum*)‡
  - Salamander (*Triturus viridescens*)‡
- Birds
  - Chicken (liver, blood)*†‡
- Mammals
  - Tree shrew‡
  - Armadillo‡
  - Hedge hog‡
  - Guinea pig‡
  - Rabbit‡
  - Rat (liver)*†‡
  - Mouse (liver, brain, thymus, spleen, kidney)*†‡
  - Hamster‡
  - Calf (thymus, liver, kidney)*†‡
- Primates
  - Tarsier‡
  - Slow Loris‡
  - Potto‡
  - Capuchin‡
  - Galago‡
  - Vervet‡
  - Owl monkey‡
  - Green monkey‡
  - Gibbon‡
  - Rhesus†‡
  - Baboon‡
  - Chimpanzee*‡
  - Human*†‡
- Plants
  - Rye (*Secale*)‡
  - Tobacco (*Nicotiana glauca*)‡
  - Bean (*Phaseolus vulgaris*)‡
  - Vetch (*Vicia villosa*)‡
  - Barley (*Hordeum vulgare*)*†
  - Pea (*Pisum sativum* var. Alaska)*†
  - Wheat (*Triticum aestivum*)*‡
  - Onion (*Allium* sp.)*

* Rate of reassociation measured directly by hydroxyapatite fractionation or measurement of optical hypochromicity as a function of time or both. † Labeled, sheared fragments bind to DNA from the same species embedded in agar at a C_0t so low that repetition must be present. ‡ Sheared nonradioactive fragments of DNA from the listed organism complete with the DNA-agar reaction (†) of a related species, reducing the amount of labeled DNA which binds to the embedded DNA.

Results obtained from DNA-agar experiments (*2, 3*) expand the list of higher organisms which contain repetitious DNA. The reassociation conditions of the DNA-agar technique yield a C_0t between 1 and 100. For DNA from higher organisms only repeated sequences will reassociate appreciably at these C_0t's. Therefore, the reassociation detected in DNA of higher organisms by the DNA-agar technique has been due to the reassociation of repetitious DNA. A list of organisms in which repeated DNA sequences have been found is shown in Table 1. Since so many types of organisms are represented it seems virtually certain that repetitious DNA is universally present in higher organisms. In assembling this table we have made use of several sets of results (*3; 35*).

The species of bacteria (*E. coli, Clostridium perfringens, Proteus mirabilis*) that have been examined do not contain repetitious DNA detectable by our methods. In none of these cases was reassociated DNA of low thermal stability observed. In all cases the kinetic curve for reassociation apparently contained only the one major component. As a further check for repetitious DNA in *E. coli*, the first small fraction to reassociate ($C_0t = 0.5$) was isolated on hydroxyapatite and shown to reassociate at the same rate as most of the *E. coli* DNA. While the sensitivity of the test is high, the existence of a small amount of repetitious DNA cannot be ruled out. Optical measurement of the reassociation kinetics on unfractionated DNA from several viruses (simian virus

40 and bacteriophages T4 and lambda) has likewise given no evidence of repeated DNA sequences.

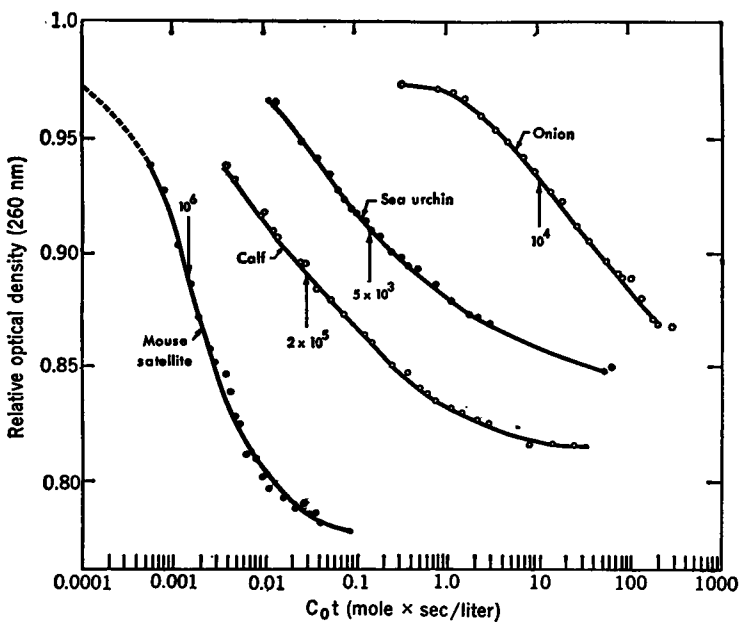
Only a very small repetitive fraction has been detected in DNA from *Saccharomyces cerevisiae*. Because of its small quantity and low native thermal stability it can tentatively be identified as mitochondrial DNA. At this moment, the relatively fragmentary evidence suggests that eukaryotes (except possibly yeast) contain repetitious DNA while prokaryotes do not. A great number of measurements will be necessary to ascertain the boundary between those life forms which do and those which do not have repetitious DNA.

Table 1 also describes interactions of DNA from a variety of tissues. We have seen no evidence for a variation in the pattern or amount of repeated sequences between different tissues of a given species or individuals of a species. In this work, sensitive tests for differences were not made; however, earlier experiments of McCarthy and Hoyer (*36*) with the DNA-agar method were specifically designed to detect variation of DNA from tissue to tissue. These may now be interpreted as showing that the repeated sequences in the mouse DNA occur to about the same extent in many tissues and in cultured cell lines.

The Precision of Repetition

When DNA strands which are not perfectly complementary reassociate, the resulting pairs have reduced stability. This effect supplies a method for measuring the degree of sequence difference among DNA strands. Measurements with artificial polymers indicate that, when 1 percent of the base pairs are not complementary, the temperature at which dissociation occurs (melting temperature, T_m) is about 1°C lower than that for perfectly complementary strands (*37, 12*). The data of Fig. 6 show that shearing bacterial DNA to small fragments, and its dissociation and reassociation, do not have a large effect on the melting temperature. This means that in the helically paired regions virtually all of the bases in reassociated bacterial DNA are properly matched. Thus the strong reduction in thermal stability observed for calf DNA indicates actual dissimilarities in the paired sequences of the reassociated DNA.

Figure 7 shows the change in adsorbance with temperature of a repetitious fraction of calf thymus DNA (prepared by hydroxyapatite fractionation of DNA sheared at 3.4 kilobars). The optical density of this fraction changes over a wide range of temperature. Most of the thermal dissociation occurs below the temperature at which native DNA begins to dissociate. The change of absorbancy shows that base-paired structure has been formed in significant amount. The broad range of dissociation, in turn, indicates imprecise pairing. This observation confirms and extends earlier measurements with DNA-agar (*11*) in which more reassociation occurred at lower temperatures of incubation both for intraspecies and interspecies pairs. The temperature and salt concentration during a reassociation incubation establish a criterion of precision in that pairs form only if they



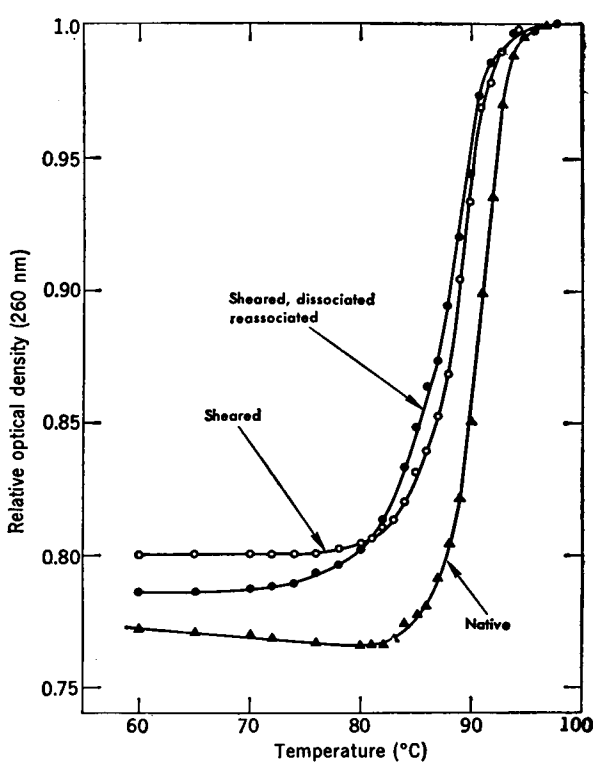


Fig. 5 (left). Optical reassociation curves of repetitious DNA fractions from various organisms. All of the fractions were purified on hydroxyapatite with only minor modifications in the procedure for each different DNA. The left scale gives the ratio of the optical density at 60°C to the initial value measured at 98°C. All of the fractions except that from onion were reassociated in 0.08*M* phosphate buffer at 60°C. The onion DNA was reassociated in 0.24*M* phosphate buffer at 60°C. The onion points were plotted a factor of 5 to the right to allow for the increased rate of reassociation, and give approximately the curve that would be observed in 0.08*M* phosphate buffer. The arrows permit estimation of the average degree of repetition in each case. They are located at the C_0t at which a fraction with the indicated degree of repetition would be half reassociated. The genome size and the amount of the rapidly reassociating fraction enter into the calculation in each case. Fig. 6 (right). Melting curves of *E. coli* DNA in 0.12*M* phosphate buffer. Open circles, native DNA sheared at 3.4 kilobars; closed circles, similarly sheared DNA dissociated (100°C, 5 minutes) and reassociated by incubation at 60°C in 0.12*M* phosphate buffer; triangles, native unsheared DNA. Shearing at 3.4 kilobars dissociates a part of the DNA, accounting for the somewhat greater hyperchromicity of the reassociated DNA.

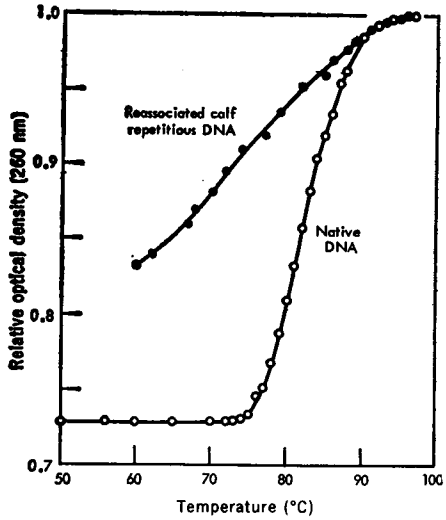


Fig. 7. DNA melting curves in 0.08*M* phosphate buffer. Open circles, unsheared native calf DNA; closed circles, reassociated calf repetitious DNA, sheared at 3.4 kilobars. In this solvent, single-stranded DNA gives an absorbancy change of only 3 percent from 60° to 95°C (see Fig. 9).

are stable above the incubation temperature. Thus the incubation temperature determines which set of sequences will reassociate and controls the resulting melting temperature.

Raising the temperature of hydroxyapatite causes adsorbed double-stranded DNA to dissociate. The resulting single-stranded fragments may then be eluted.

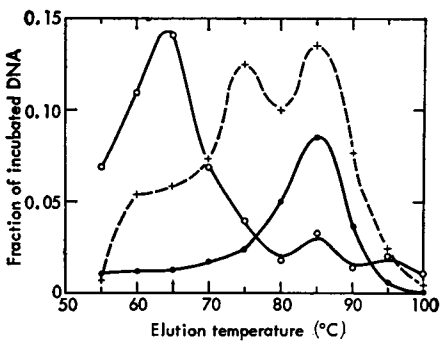


Fig. 8. Thermal fractionation on hydroxyapatite of reassociated salmon sperm DNA. DNA sheared at 3.4 kilobars was incubated at 50°C in 0.14*M* phosphate buffer (C_0t, 370) and passed over hydroxyapatite at 50°C in 0.14*M* phosphate buffer. The adsorbed DNA was eluted by exhaustive washing (0.14*M* phosphate buffer) at intervals of 5°C (dashed line and crosses). To show specificity, four fractions (65°, 70°, 85°, 90°C) were again denatured (100°C, 5 minutes) and reincubated (50°C, 0.14*M* phosphate buffer, C_0t about 10) and readsorbed on hydroxyapatite at 50°C. Two of these were again thermally eluted from a column: open circles, 65°C fraction; closed circles, 85°C fraction. The other two were eluted with 0.4*M* phosphate buffer and melted in the spectrophotometer as shown in Fig. 9.

When dissociation is plotted against temperature, the profiles are very similar to those measured by change in ultraviolet absorbancy in free solution (*15, 22*).

Figure 8 shows reassociated repetitive salmon DNA fractionated with hydroxyapatite on the basis of its thermal stability. Sheared DNA from salmon was dissociated and then incubated at 50°C in 0.14*M* phosphate buffer (*31*) for a C_0t of 270, and the reassociated DNA was adsorbed on hydroxyapatite at 50°C, 0.14*M* phosphate buffer. The temperature of the column was raised in steps of 5°C, and at each temperature the dissociated DNA was washed from the column with 0.12*M* phosphate buffer. The resulting chromatogram (Fig. 8, dashed line) shows a broad range of thermal stability. In order to establish the specificity of the fractionation, samples eluted at 65° and 85°C were incubated again at 50°C. They were then readsorbed and reanalyzed as before.

The strand pairs formed during the second incubation are ordinarily not the ones that were originally eluted. Instead, they are new duplexes formed by randomly assorted pairings among the selected set of strands. In each case, however, the same average degree of precision of relationship results. The portion eluting at 65°C shows a peak again at 65°C, and the 85°C portion peaks at 85°C. The degrees of sequence divergence are thus characteristic of these sets of fragments. Similar studies have been done with calf thymus DNA with entirely comparable results. In addition, experiments with labeled calf DNA fractions indicate that little sequence homology exists between precisely and imprecisely reassociating sets of repetitive DNA.

Length of Repeated Sequences

Are reassociated repeated sequences complementary only in short regions or are they complementary over most of their length? The thermal stability of a pair does not by itself answer this question since it appears that complementary sequences, 100 nucleotide pairs long, will have a thermal stability approaching that of very long complementary sequences (*38, 21*). However, ultraviolet hyperchromicity is a measure of the extent of sequence matching. Therefore, the hyperchromicity of a preparation of strand pairs gives a measure of the fraction of the total

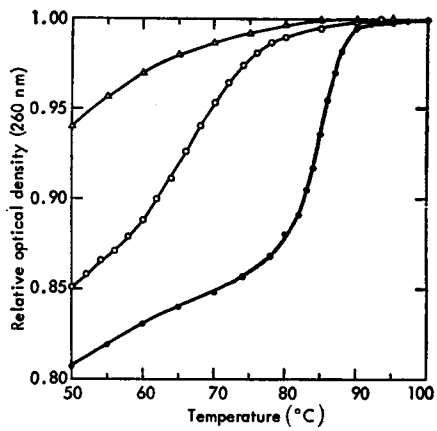


Fig. 9. Spectrophotometric melting curves in 0.14*M* phosphate buffer of fractions of salmon sperm DNA. Fractions were prepared as described in Fig. 8: closed circles, fraction eluted at 90°C; open circles, fraction eluted at 70°C. The upper curve (open triangles) is for the DNA which did not bind to hydroxyapatite (50°C, 0.14*M* phosphate buffer) in the first incubation and is therefore purely single stranded.

length which is complementary. Results for two such preparations are shown in Fig. 9. Native, completely complementary salmon DNA has a hyperchromicity of about 0.25. Single-stranded DNA has a hyperchromicity of 0.06 and melts mostly at lower temperatures, as shown by the top curve.

The 70° and 90°C fractions each have about half the hyperchromicity of native DNA. From this we may conclude in each case that the base-paired regions of the reassociated repetitious DNA include about half of the nucleotides of the fragments.

Several complicating factors interfere with a more firm conclusion. Reassociated fragments will, in general, have single-stranded ends, since two complementary fragments rarely terminate at the same points in the sequence. All degrees of overlap will occur, and for first pairing the expected hyperchromicity is between one-half and two-thirds that for native DNA. We do not know the further extent of pair formation involving the single-stranded ends in these preparations. Finally, of course, repeated sequences have diverged from each other, and the unmatched nucleotides occurring within the paired sequences reduce the hyperchromicity.

These measurements are corroborated by the hyperchromicity we have observed for reassociated repetitive DNA fractions from many organisms. It usually falls between 0.10 and 0.20. Some examples are shown on Fig. 5. The few CsCl equilibrium centrifuga-

tion measurements we have made show a marked decrease in density upon reassociation which also implies a good extent of complementary pairing of repetitive DNA.

It appears that, on the average, repeated sequences are not extremely short (not less than 200 nucleotides) and may be much longer than our fragments, which average perhaps 400 nucleotides. In other words, wherever a region of sequence homology occurs between two fragments of the genome, it is likely to continue for at least several hundred nucleotides. It does not usually continue perfectly, however, since the reduced thermal stability observed implies that local interruptions of the homology must be scattered through the regions of homology.

Nonrepeated DNA of Higher Organisms

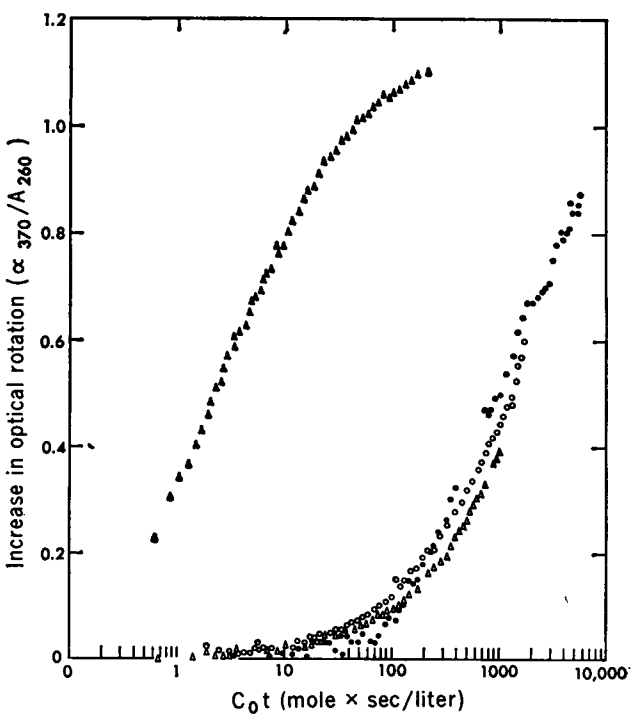
Somewhat more than half of the DNA from mouse or calf may be recovered in the single-stranded state after an extensive incubation (Fig. 3; $C_0t = 100$). The subsequent, very slow, reassociation of this fraction has been measured by three methods—hydroxyapatite adsorption (Fig. 3), optical rotation increase (370 nanometers) (Fig. 10), and hypochromicity (260 nanometers). The measurements establish that this fraction reassociates accurately and nearly completely. Experiments (*15*) indicate species dependence and therefore sequence specificity of the reassociation of the very slow fraction.

The curves shown on Fig. 3 give the results of a measurement of the rate of reassociation of the slow fraction of calf DNA in comparison with that of *E. coli* DNA. Labeled *E. coli* DNA was present with the calf DNA during the incubation as an "internal standard" (*33*). Thus these measurements yield a relatively precise measure of the concentration of complementary sequences in the *E. coli* DNA compared to those in the calf DNA.

The DNA content of the bull sperm is 3.2×10^9 nucleotide pairs per cell (*30*), and Cairn's measurement (*26*) gives 4.5×10^6 nucleotide pairs for the size of the *E. coli* genome. The ratio of these numbers is 710, and the ratio of the C_0t for half reaction of the slow part of the calf DNA curve to that for *E. coli* is 690. This establishes with greater accuracy the conclusion drawn from Fig. 2 that under these conditions about 60 percent of the calf DNA does not exhibit repeated sequences. If the slowly reassociating sequences were nonrepeating (unique), they should also reassociate accurately since only the precisely complementary strand will be present. In other words, both the hyperchromicity and thermal stability of the reassociated pairs should approach that of native DNA of this fragment size.

Figure 11 shows the hyperchromic melting curves for very rapidly and very slowly reassociating fractions of calf thymus DNA. The slowly reassociating fraction was prepared by repeated incubation and passage through hydroxyapatite ($C_0t = 100$, 60°C, 0.12*M* phosphate buffer). It was then incubated extensively ($C_0t = 1000$, 60°C, 0.24*M* phosphate buffer), and the reassociated portion was isolated by binding to hydroxyapatite. The C_0t was sufficient only to reassociate about half of the DNA strands. Thus, little concatenation (*12*) had occurred, and the hyperchromicity had not achieved its final value. Nevertheless, the nonrepetitious fraction has almost 75 percent of the hyperchromicity of native high-molecular-weight DNA and a relatively sharp thermal transition. Much of the difference between this curve and that for native DNA (Fig. 7) is due to the fact that the DNA had been sheared before reassociation. Native high-molecular-weight and sheared *E. coli* DNA (Fig. 6) differ to about the same extent.

The rapidly reassociating fraction was eluted from the hydroxyapatite in a reassociated state and was diluted to the proper solvent and concentration for the measurement. A very brief incubation was used in this case, and the melting curve is strongly influenced by



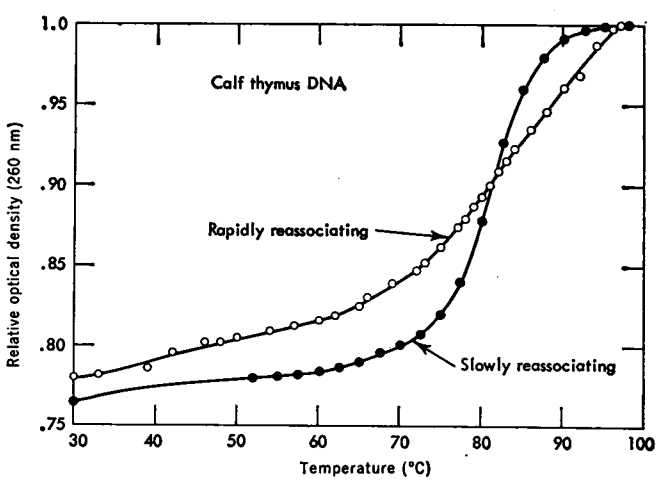


Fig. 10 (left). Measurement by optical rotation at 370 nanometers (Rudolph recording spectropolarimeter) of the reassociation of *E. coli* DNA and the slowly reassociating fractions of mouse and calf DNA (0.24*M* phosphate buffer, 60°C). Closed triangles, *E. coli* DNA at 0.69 mg/ml in a 10-cm cell. Open triangles, mouse DNA fractionated to remove rapidly reassociating sequences. Open circles, calf thymus DNA fractionated to remove rapidly reassociating sequences. Closed circles, a second, similar, preparation of calf thymus DNA. The reaction rates shown here are nearly threefold greater than those shown in Fig. 2 because of the greater salt concentration. Fig. 11 (right). Hyperchromic melting curves for rapidly reassociating and slowly reassociating fractions of calf thymus DNA in 0.08*M* phosphate buffer. Fractions were prepared with hydroxyapatite and reassociated as described in the text.

a particular fraction (*39*) which reassociates very rapidly, and apparently has a high guanine-cytosine content and melting temperature. A more typical melting curve of a repetitive DNA fraction prepared after a more extensive incubation is shown on Fig. 7.

Reassociation of unique sequences of calf DNA was originally observed in our laboratory by means of the spectropolarimeter. This instrument is particularly suitable since a high concentration of DNA is required both for the large C_0t and for an accurate measurement of the change in optical activity. The slowly reassociating DNA was prepared for this purpose from commercial calf thymus DNA by hydroxyapatite fractionation and lyophilization.

In the region of the spectrum above 300 nanometers there is a good contrast in specific rotation between the native and denaturated states, and little light is absorbed by the DNA. The reassociation of the purified slow fraction of both mouse and calf thymus DNA occurs at about 1/500 the rate of that of *E. coli* DNA under the same conditions (Fig. 10). Within the accuracy of the measurements this is the expected rate for a 60 percent fraction of mammalian DNA which has no repeated sequences. The reassociation of the unique sequences of mammalian DNA is therefore confirmed by measurement of the change in optical rotation.

Amphiuma is the organism with the largest known genome, having 8×10^{10} nucleotide pairs of DNA per haploid cell. The expected C_0t for half reassociation in the absence of repetitions would be 80,000 under the conditions of Fig. 2. *Amphiuma* DNA was fractionated on hydroxyapatite after shearing and incubation (C_0t = 80, 50°C, 0.14*M* phosphate buffer). Only 20 percent of the DNA was recovered in the slowly reassociating fraction. This fraction was incubated (60°C, 0.24*M* phosphate buffer, 5 milligrams per milliliter), and samples were analyzed on hydroxyapatite every few days. It exhibited the slowest reassociation reaction that has been observed, reaching half reassociation with a C_0t of 20,000. This apparent agreement with the expected rate cannot be considered definitive until a number of controls are done. Nevertheless, it appears likely that a significant fraction of the genome of even this creature is made up of unique sequences.

Patterns of Repetition

The rate of reassociation of the DNA of one organism can be evaluated over the whole course of the reaction (C_0t from 10^{-4} to 10^4). Individual measurements of reassociation at several concentrations are required, and fractionation of the DNA is useful. From these measurements the amount of DNA with various degrees of repetition may be calculated. The result is a repetition-frequency spectrogram for the DNA of the particular organism, such as the tentative one for mouse DNA shown in Fig. 12. This curve is correct in its broad aspects but has some indefiniteness in detail. The width of the peaks results in part from the difficulty of resolving reassociation rates that differ by less than a factor of 10.

The large peak at the left on Fig. 12 is due to the mouse satellite DNA (*13*). Such a class of DNA molecules which can reassociate with each other is called a family since the similarity in sequence implies a common origin. Correspondingly, other classes capable of reassociating are called families even though the precision of reassociation is less, presumably due to divergence of the members since the formation of the family.

Repetition-frequency spectra could also be derived for calf DNA and salmon DNA from Figs. 3 and 4. However, they would differ from that of Fig. 12. Neither would show the large isolated peak of 10^6 copies. The calf DNA would show a large broad peak in the region of 10^4 to 10^5 copies (40 percent of the DNA), little if any DNA with a small degree of repetition, and of course a large peak of unique DNA.

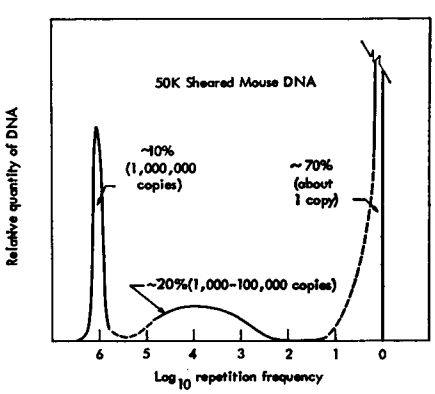


Fig. 12. Spectrogram of the frequency of repetition of nucleotide sequences in the DNA of the mouse. Relative quantity of DNA plotted against the logarithm of the repetition frequency. These data are derived from measurements of the quantity and rate of reassociation of fractions separated on hydroxyapatite. The dashed segments of the curve represent regions of considerable uncertainty.

What length of DNA sequence has been replicated to form families of repeated sequences? Forty percent of the calf DNA behaves as repeating DNA. The total quantity of repeated sequences is 1.3×10^9 nucleotide pairs per cell. Lengths of DNA totaling 13,000 nucleotide pairs copied 100,000 times would have about the same total quantity and average repetition frequency as the repetitious DNA of the calf. Such a homogeneous set of fragments would have the smallest possible information content that could be present in the repeated DNA of the calf. The situation is known to be more complex, however, and the potential information content of the repetitious DNA fraction is very much greater for the following reasons. (i) The nucleotide sequences of the members of a typical family are similar to each other but not identical. The differences may be of great genetic significance. (ii) A small amount of DNA probably occurs in families made up of long sequences repeated only a few times. (iii) The repeated sequences or fragments of them have been translocated into various parts of the genome, and their location and relationship to their neighbor sequences may be important.

In this regard an observation of Britten and Waring (*12*) with higher-organism DNA is significant. When 5 to 10 million dalton DNA strands are dissociated and incubated (C_0t = 1), large particles form creating a visible haze in solution, and most of the DNA may be sedimented to the bottom of a tube after centrifugation for 5 minutes at 10,000*g*. Apparently some regions of most of these DNA fragments are members of families of repeated sequences. Reassociation of the repeated sequences links the fragments into a large network. A number of such measurements with various fragment sizes indicate that repeated sequences are scattered throughout the genome. This extensive interspersion of members of families of repeated sequences may be related to their function. They could, for example, have regulatory or structural roles which would lead to such a distribution. This dispersion may, however, simply represent the degree of translocation of sequence fragments that has occurred during the evolution of higher organisms.

The Criteria for Repeated Sequences

The presence in a DNA solution of a fraction which reassociates rapidly indicates that certain nucleotide sequences are present at a higher concentration than the remainder. If the DNA's were derived from a single organism we can usually conclude that these sequences recur repeatedly in its DNA. The conclusion is essentially certain if the reassociation exhibits the variation of rate with concentration of a second-order collision-controlled reaction.

If such a concentration dependence is not shown other possibilities arise. For example, the rapidly reassociating DNA's could be in closed circular form, the two strands could be cross-linked, or a sequence might occur which contained its own complement and could renature by folding. An example of the latter is the satellite from crab DNA (*40*) which is principally an alternating sequence of adenine and thymine. In all of these cases, a bimolecular collision is not involved, and the reaction is extremely rapid under optimum conditions for reassociation. With the methods used in this work the rate would be so fast that it could not be observed. Thus, under our usual conditions the observation of a measurable rate of reassociation (faster than that expected from the genome size of the organism) is an almost certain indication of the presence of repeated sequences.

What are the limits of accuracy in the calculation of repetition frequency from a measured rate of DNA reassociation? If only a part of the length of fragments are complementary to each other the rate of reassociation will be reduced. The fraction of the nucleotides which are complementary in typical reassociated pairs of fragments containing repeated sequences is not well known but some limits can be set. The hyperchromicity observed for reassociated repetitive DNA ranges from just less than half up to nearly that characteristic for reassociated bacterial or viral DNA. This evidence implies that half or more of the nucleotides are complementary in typical reassociated repetitive DNA. We believe that under these conditions the reduction in the rate of reassociation is not large. However, the frequency of repetition may be somewhat underestimated in this work.

There is a possibility also of a potential overestimate in the quantity of repeated DNA sequences. When reassociation is assayed with hydroxyapatite, all strands of DNA which contain a sizable double-stranded region will bind. The minimum double-stranded region which will adsorb to hydroxyapatite under the conditions used is not known but is much smaller than the fragment size used. Thus, a certain fraction of the nonrepetitive DNA will be included in the measured repetitive fraction. If the repeated sequences occur in stretches which are long compared with the fragment size, this error will be small. Partly for this reason small fragments are used in this work.

There is evidence that homopolymer clusters occur in DNA (*41*). It has not yet been demonstrated that such clusters influence DNA reassociation. The quantity present in the DNA of higher organisms is not known. Nevertheless, in the following paragraphs we attempt an estimate of the maximum effect homopolymer clusters could have on the rate of DNA reassociation.

If the homopolymer clusters were long enough to form a large fraction of the length of a set of DNA fragments they would simply form an extreme example of repeated sequences. The reassociation rate constant would be that shown for the polyuridylic acid and polyadenylic acid pair on Fig. 2. The reassociation would appear to be instantaneous under our conditions, except with very low concentrations of DNA.

Very short homopolymer clusters might play two possible roles. They might cause fragments not complementary over the rest of their length to form stable structures which are paired only in the cluster region. These would form a class (not yet observed) of repetitive DNA with very low hypochromicity. Short homopolymer clusters might also increase the rate of nucleation of fragments which were fully complementary and thus increase their rate of reassociation.

The limit for the maximum possible factor of increase is less than twice the number of nucleotides in the homopolymer cluster. This can be seen by making the most favorable assumptions. (i) Nucleations in the homopolymer regions of otherwise noncomplementary fragments do not interfere (that is, they dissociate quickly); (ii) all nucleations occur independently and the rate of reassociation is proportional to the number of "in register" collisions possible; (iii) all possible "registrations" of the homopolymer cluster (twice the number of nucleotides in the cluster) lead to reassociation of major complementary regions if present.

Since condition (i) would not be met for homopolymer clusters longer than 20, the factor of increase in rate must be less than 40. Since condition (iii) is not likely to be met, the factor of increase is probably much less than that. This factor is not large enough to affect the conclusion about the general occurrence of repeated sequences.

Implications of Repeated Sequences

These studies have revealed new properties of the DNA of higher organisms which must be attributed to the repetition of nucleotide sequences. Some of these properties are summarized in Table 2. In general, more than one-third of the DNA of higher organisms is made up of sequences which recur anywhere from a thousand to a million times per cell. Thus the genetic material is not a collection of different and unrelated genes. A large part is made up of families of sequences in which the similarity must be attributed to common origin.

A minor degree of sequence repetition is to be expected from studies of protein sequences (*42*). The hemoglobin group shows similarities in sequences, and these similarities point to common origin of part or all of their structural genes. Trypsin and chymotrypsin also show similarities. There is evidence that, in some cases, different segments of the amino acid sequence of a given protein may have arisen by duplication and insertion of an earlier short segment. In addition to genetic evidence, banding patterns in polytene chromosomes show that gene duplication occurs (*43*). The genome sizes (*25*) of higher organisms range from 10^8 to 10^{11} nucleotide pairs (*30*). There is no doubt that a great increase in DNA content has occurred during the evolution of certain species.

These observations suggest that a degree of nucleotide sequence repetition might be observed in the DNA. It must be emphasized that they do not imply that DNA sequence repetition occurs on anything approaching the scale reported here. The very large number of members in the families of repeated sequences remains a most surprising feature for which an explanation must be sought. It may be reasonably predicted

Table 2. Characteristics of DNA reassociation.

Nonrepetitive DNA	Repetitive DNA
Source	
Bacteria Viruses	Vertebrates Invertebrates Higher plants Euglena Dinoflagellate
Rate of reassociation	
One rate, inversely proportional to DNA content per cell or particle	Many different rates. Slowest inversely proportional to DNA content per haploid cell. Fastest up to 10^6 times faster
Extent of reassociation	
Excellent, up to 90 percent reformed helices (no strong effect of fragment size)	Good if DNA cut into small fragments. Poor if DNA is of high molecular weight
Stability of reassociated DNA	
Temperature at which strands separate (T_m) almost equal to that of native DNA	Some with T_m near that native DNA and many lower degrees of stability
Particle size of reassociated DNA	
Several times the fragment size due to pairing of free single-stranded ends (concatenation)	Enormous, if DNA fragments are large, due to multiple interconnections (network formation)

that large-scale new patterns of relationship among the proteins await discovery.

Certain minor classes of DNA probably consist of many copies of a short sequence. It appears likely that there are hundreds or thousands of similar ribosomal genes (*44*) and in certain cells, at least, thousands of similar, if not identical, copies of mitochondrial DNA (*45*). Taken together, such classes of DNA do not add up to more than a percent of the total DNA and, compared to the bulk of the repeated sequences, have a relatively low repetition frequency.

If many DNA sequences in a chromosome are similar to each other and adjacent, high rates of unequal crossing-over might occur. Although there is genetic evidence (*43*) that this occurs, it has not been considered common. Presumably higher organisms are protected from the lethal genetic events (*46*) that the families of repeated sequences might induce.

There is a certain amount of evidence that repeated sequences are genetically expressed. Pulse-labeled RNA (presumptive messenger) has been hybridized with the DNA of higher organisms (*47*). In most of these studies, hybrids were only observed with RNA that was complementary to families of repeated sequences. Due to the small C_0t, hybrids between RNA and nonrepeated DNA sequences of higher organisms apparently did not occur.

The RNA populations made from repeated DNA sequences may have some role (perhaps regulatory) other than as messengers carrying structural information for protein synthesis. However, this is an unlikely (and certainly an unpopular) possibility. A good working hypothesis is that repeated sequences commonly occur in structural genes. In any case, transcription as complementary RNA is direct evidence for the genetic function of at least some of the repeated DNA sequences. In the course of embryonic development and during liver regeneration (*47*), changes occur in the pattern of types of hybridizable pulse-labeled RNA. These results suggest that during the course of differentiation different families of repeated sequences are expressed at different stages.

Origin and Age

The families of repeated sequences range from groups of almost identical copies (for example, mouse satellite DNA) to groups with sufficient diversity that, after reassociation, only structures of low stability are formed among the members. It seems likely that this situation has arisen from large-scale precise duplication of selected sequences, with subsequent divergence caused by mutation and the translocation of segments of certain member sequences. We cannot now describe the history of growth and divergence of any particular family of repeated sequences. However, the few measured properties of the repetitive DNA permit some inferences.

The extensive studies of Hoyer *et al.* (*3, 10, 11*) supply a measure of the repeated sequences held in common among different species. Because of the small C_0t used in their work only the reassociation of DNA sequences repeated in each organism was observed. These measurements were carried out at various temperatures (*11*), and the results were correlated with the period of time after divergence of the lines leading to the modern species (*10*).

These data show a low average melting temperature if strands of DNA from different species are reassociated. The longer the period after divergence of the species, the greater the reduction in thermal stability. This evidence indicates that the members of families of repeated sequences in the DNA of a species slowly change in nucleotide sequence.

It is an unlikely possibility that all the members of a family of repeated sequences in one organism undergo the same changes. This would involve either very severe selection on all the members or a complex event such as discarding all but one of the members of a family and then multiplying the remaining member 10,000 or 100,000 times. The much more appealing and simpler model is that the nucleotide sequences of the members of the families are not conserved by severe selection. The members may then change slowly and independently of each other leading, after a long period of time, to families with widely divergent members such as are observed.

In addition to the divergence of preexisting families, new families are produced in each species. Analysis of the present data on repetition frequency distribution (*14*) suggests that they result from relatively sudden events which we have called saltatory replications. Figure 13 symbolizes the resulting view of the history of families of repeated sequences. Along one axis is the time since formation of the family. Along the other axis is the temperature at which duplexes among the family members dissociate. The third axis represents the number of members. Thus the area of one of the peaks indicates the number of members of the family. The temperature at which the peak occurs is a measure of the extent of sequence difference among the members. The diagram is not intended to be quantitative, although we have used the estimates that are available for vertebrates. The frequency of events and rates of divergence are also probably very different for other phyla.

Even if saltatory replications are as

rare as indicated on Fig. 13, certain stages of the process may be relatively common. We know nothing of the mechanism of the process, but the following steps seem necessary. (i) A sequence undergoes manyfold replication; (ii) the copies are integrated into the chromosome; (iii) they become associated with a favorable genetic element, and (iv) they are disseminated through the species by natural selection.

Each of the succeeding stages is likely to have a very low probability of occurrence, and thus the actual event of manyfold replication may occur fairly commonly and, in principle, be observable in individual organisms, in analogy to a somatic mutation. It does not seem impossible that, some time in the future, saltations may be artificially introduced into populations as mutations can already be.

Speculation on Their Function

A concept that is repugnant to us is that about half of the DNA of higher organisms is trivial or permanently inert (on an evolutionary time scale). Furthermore, at least some of the members of DNA families find expression as RNA. We therefore believe that the organization of DNA into families of related sequences will ultimately be found important to the phenotype. However, at present we can only speculate on the actual role of the repeated sequences.

Multiple, nearly exact copies of a gene could provide higher rates of synthesis. This might be true for structural proteins required in large amounts and is very likely true for ribosomal RNA. Multiple similar copies could provide a class of similar protein chains as appear to occur in antibody proteins (*48*). However, their role could not be limited to the immune system since they occur in large quantities in the plants and other organisms in which antibodies have not been observed.

The DNA of each vertebrate that has been examined contains some families with 100,000 members or more. This very large number suggests a structural (*49*) or regulatory role. However, the significance of the very large number might be less direct. It might, for example, raise to a useful level the probability of some rare event such as the translocation of certain DNA sequence fragments into adjacent locations in the genome.

Saltatory replications of genes or gene fragments occurring at infrequent intervals during geologic history might have profound and perhaps delayed results on the course of evolution. In the following quotation Simpson (*50*) raises some relevant questions with regard to evolutionary history.

> The history of life is decidedly nonrandom. This is evident in many features of the record, including such points already discussed as the phenomena of relays and of major replacements at defined times. It is, however, still more striking in two other phenomena copiously documented by fossils. Both have to do with evolutionary trends: first, that the direction of morphological (hence also functional and behavioral) change in a given lineage often continues without significant deviation for long periods of time and, second, that similar or parallel trends often appear either simultaneously or successively in numerous different, usually related, lineages. These phenomena are far from universal; they are not "laws" of evolution; but they are so common and so thoroughly established by concrete evidence that they demand a definite, effective directional force among the evolutionary processes. They rule out any theory of purely random evolution such as the rather naive mutationism that had considerable support earlier in the twentieth century. What directional forces the data do demand, or permit, is one of the most important questions to be asked of the fossil record.

The appearance in a genome of many thousands of copies of a gene could have evolutionary significance. Perhaps not many copies would be actually expressed. Mutation, translocation, and recombination with other genes might yield new genetic potential. If the early effects were selectively advantageous, the repeated DNA sequences could be introduced into the population. The dynamics of selection for this set of genes would be fundamentally altered. Owing to the great multiplicity of copies, their selective elimination might be impossible.

Summary

The rate of reassociation of the complementary strands of DNA of viral and bacterial origin is inversely proportional to the (haploid) DNA content per cell. However, a large fraction of

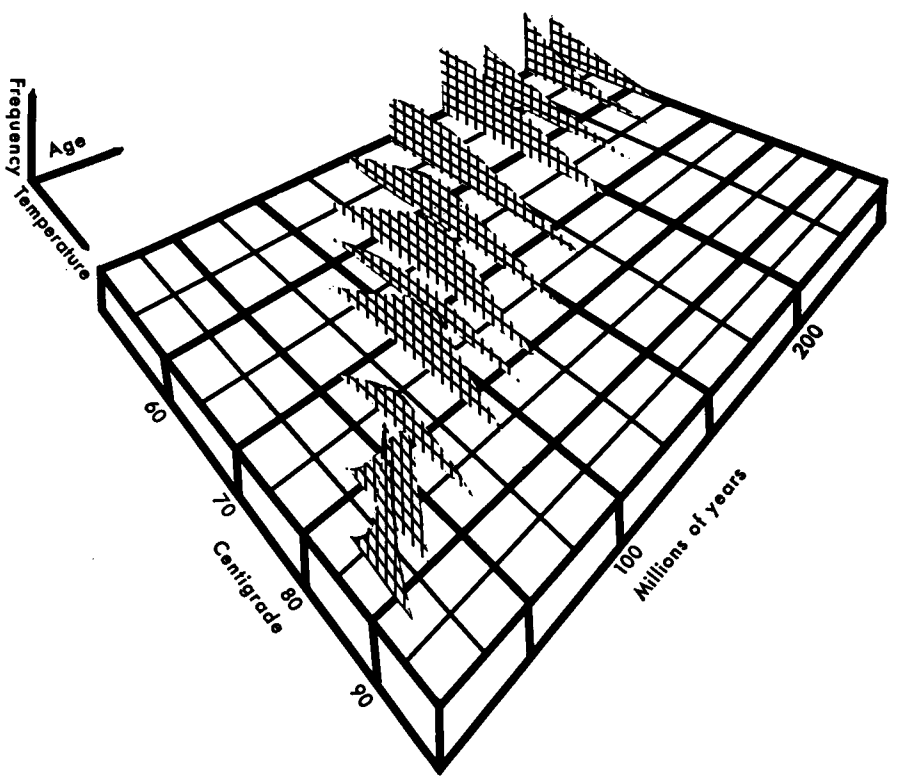


Fig. 13. A schematic diagram intended to suggest the history of the families of repeated sequences now present in the DNA of a modern creature. Each family is supposed to have originated in a sudden event (saltatory replication) at a time in the past shown on the right-hand scale. In the ensuing time, increasing divergence has occurred between the nucleotide sequences of the members of each family of repeated sequences. This divergence is represented on the left-hand scale by the thermal stability of reassociated pairs of DNA strands formed between members of each family. The height of the cross-hatched areas indicates the amount of DNA of a given thermal stability in a family of a particular age. Only a few of a potentially large number of families are indicated. The actual rate of divergence has not yet been well measured.

the DNA of higher organisms reassociates much more rapidly than would be predicted from the DNA content of each cell. Another fraction appears to reassociate at the expected rate. It is concluded that certain segments of the DNA are repeated hundreds of thousands of times. A survey of a number of species indicates that repeated sequences occur widely and probably universally in the DNA of higher organisms.

The repeated sequences have been separated from the remaining (unique sequence) DNA, and their physical properties have been studied. The range of frequency of repetition is very wide, and there are many degrees of precision of repetition in the DNA of individual organisms. During evolution the repeated DNA sequences apparently change slowly and thus diverge from each other. There appears to be some mechanism which, from time to time, extensively reduplicates certain segments of DNA, replenishing the redundancy.

References and Notes

1. J. Marmur, R. Rownd, C. L. Schildkraut, *Progr. Nucleic Acid Res.* **1**, 231 (1963).
2. B. J. McCarthy and E. T. Bolton, *Proc. Nat. Acad. Sci. U.S.* **50**, 156 (1963).
3. E. T. Bolton *et al.*, *Carnegie Inst. Wash. Year Book* **63**, 366 (1964); *ibid.* **62**, 303 (1963); B. H. Hoyer, B. J. McCarthy, E. T. Bolton, *Science* **144**, 959 (1964).
4. R. C. Warner, *J. Biol. Chem.* **229**, 711 (1957); A. Rich and D. Davies, *J. Amer. Chem. Soc.* **78**, 3548 (1956).
5. J. Marmur and D. Lane, *Proc. Nat. Acad. Sci. U.S.* **46**, 456 (1960).
6. B. D. Hall and S. Spiegelman, *ibid.* **47**, 137 (1961).
7. E. K. F. Bautz and B. D. Hall, *ibid.* **48**, 400 (1962).
8. E. T. Bolton and B. J. McCarthy, *ibid.*, p. 1390.
9. A. P. Nygaard and B. D. Hall, *J. Mol. Biol.* **9**, 125 (1964); D. Gillespie and S. Spiegelman, *ibid.* **12**, 829 (1965).
10. B. H. Hoyer, E. T. Bolton, B. J. McCarthy, R. B. Roberts, *Carnegie Inst. Wash. Year Book* **63**, 394 (1964); in *Evolving Genes and Proteins*, V. Bryson and H. Vogel, Eds. (Academic Press, New York, 1965), p. 581.
11. M. Martin and B. Hoyer, *Biochemistry* **5**, 2706 (1966); P. M. B. Walker and A. McLaren, *J. Mol. Biol.* **12**, 394 (1965).
12. R. J. Britten and M. Waring, *Carnegie Inst. Wash. Year Book* **64**, 316 (1965).
13. M. Waring and R. J. Britten, *Science* **154**, 791 (1966).
14. R. J. Britten and D. E. Kohne, *Carnegie Inst. Wash. Year Book* **66**, 73 (1967).
15. ———, *ibid.* **65**, 73 (1966).
16. J. A. Subirana and P. Doty, *Biopolymers* **4**, 171 (1966); J. A. Subirana, *ibid.*, p. 189; J. G. Wetmur and N. Davidson, *J. Mol. Biol.* **31**, 349 (1968); K. J. Thrower and A. R. Peacocke, *Biochim. Biophys. Acta* **119**, 652 (1966).
17. J. G. Wetmur, thesis, California Institute of Technology, Pasadena (1967).
18. The DNA used in this work has been sheared to a relatively uniform population of small fragments (about 400 or 500 nucleotides long) by passing it twice through a needle valve with a pressure drop of 3.4 kilobars. A specially built air-operated plunger pump was used to develop this high pressure. The DNA is denatured when sheared at 3.4 kilobars, unless a very high salt concentration is present to raise the temperature of melting (T_m). These small fragments give reproducible rates of reassociation and do not, under the usual conditions for reassociation, form large aggregates or networks.
19. For some purposes, correction can be made for different values of the parameters. There have been some quantitative measurements of the effect of temperature (*1*, *15*) salt concentration (*15–17*) and fragment size (*16*, *17*).
20. D. T. Denhardt, *Biochem. Biophys. Res. Commun.* **23**, 641 (1966).
21. B. J. McCarthy, *Bacteriol. Rev.* **31**, 215 (1967).
22. G. Bernardi, *Nature* **206**, 779 (1965); P. M. B. Walker and A. McLaren, *ibid.* **208**, 1175 (1965); Y. Miyazawa and C. A. Thomas, Jr., *J. Mol. Biol.* **11**, 223 (1965).
23. Here, *Cot* is used as a noun and may be pronounced as the homonym of "cot." A *Cot* of 1 mole × second/liter) results if DNA is incubated for 1 hour at a concentration of 83 μg/ml, which corresponds to an optical density of about 2.0 at 260 nanometers.
24. In our experience, the reassociation of purified sheared (*18*) DNA shows the concentration dependence expected for a second-order reaction. For DNA without repeated sequences, the time course also approximately follows second-order kinetics. While earlier measurements have suggested greater complexity, this is not supported by more recent work (*17*).
25. The word genome customarily means the genetic constitution of an organism. Here the genome size is taken to mean the haploid DNA content of a cell or virus particle. The number of different fragments will only be proportional to the genome size in the absence of repetition or unrecognized polyploidy.
26. J. Cairns, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 43 (1963). Because of its apparent lack of repetition and measured genome size *E. coli* DNA is used as a reference for the comparison of reassociation rates of other DNA. The use of this reference in future work will permit comparison of rates where fragment size and other conditions affecting absolute rate may vary.
27. I. Rubenstein, C. A. Thomas, A. D. Hershey, *Proc. Nat. Acad. Sci. U.S.* **47**, 1113 (1961); J. Cairns, *J. Mol. Biol.* **3**, 756 (1961); E. Burgi and A. D. Hershey, *ibid.*, p. 458.
28. M. A. Billeter, C. Weissmann, R. C. Warner, *J. Mol. Biol.* **17**, 145 (1966).
29. P. D. Ross and J. M. Sturtevant, *J. Amer. Chem. Soc.* **84**, 4503 (1962).
30. B. J. McCarthy, in *Progr. Nucleic Acid Res. Mol. Biol.* **4**, 129 (1965); T. Mann, in *The Biochemistry of Semen and of the Male Reproduction Tract* (Wiley, New York, 1964), p. 147.
31. This phosphate buffer is composed of an equimolar mixture of Na_2HPO_4 and NaH_2PO_4. The indicated molarity is for the phosphate. The sodium-ion concentration is 1.5 times greater.
32. D. E. Kohne and R. Britten, in preparation.
33. The variables controlled in this way are salt concentration, temperature, and viscosity. In addition, any possible nonspecific interactions in the DNA at this high concentration will have similar effects on both the *E. coli* and calf DNA reassociation reactions. The half reaction *Cot* for *E. coli* DNA in Fig. 3 is 6.0, whereas on Fig. 2 it is 8.0. Reactions usually appear twofold faster when assayed with the hydroxyapatite method as compared to the optical method since the fraction of fragments reassociated is measured in one case, while the fraction of total strand length reassociated is measured in the other case (*15*). On Fig. 3 there may be a 50 percent increase in the *Cot* for half reaction for the data taken at 8600 μg/ml. This decrease in the rate of reassociation is due to the increased viscosity of the incubation solution.
34. The DNA was prepared by a combination of methods of J. Marmur, *J. Mol. Biol.* **3**, 208 (1961); K. I. Berns and C. A. Thomas, *Biophys. Soc. Abstr.* (1964); B. J. McCarthy and B. H. Hoyer, *Proc. Nat. Acad. Sci. U.S.* **52**, 914 (1964). Purity was tested in the spectrophotometer by melting in 0.12*M* phosphate buffer. We required that there be no measurable rise in optical density between 40° and 70°C, and that a normal melting curve was obtained with a hyperchromicity of at least 25 percent of the absorbancy at 98°C. Commercial calf and salmon DNA were utilized in some experiments, and no differences were observed with results obtained with DNA prepared from fresh tissue.
35. B. J. McCarthy and E. T. Bolton, *Proc. Nat. Acad. Sci. U.S.* **50**, 156 (1963); B. H. Hoyer, B. J. McCarthy, E. T. Bolton, *Science* **144**, 959 (1964); E. T. Bolton *et al.*, *Carnegie Inst. Wash. Year Book* **64**, 314 (1965); H. Denis, *ibid.*, p. 455.
36. B. J. McCarthy and B. H. Hoyer, *Proc. Nat. Acad. Sci. U.S.* **52**, 915 (1964).
37. E. K. F. Bautz and F. A. Bautz, *ibid.*, p. 1476; T. Kotaka and R. L. Baldwin, *J. Mol. Biol.* **9**, 323 (1964).
38. A. Rich, *Proc. Nat. Acad. Sci. U.S.* **46**, 1044 (1960); M. N. Lipsett, L. A. Heppel, D. F. Bradley, *J. Biol. Chem.* **236**, 857 (1961); M. N. Lipsett, *ibid.* **239**, 1256 (1964).
39. E. Polli, G. Corneo, E. Ginelli, P. Bianchi, *Biochim. Biophys. Acta* **103**, 672 (1965).
40. N. Sueoka, *J. Mol. Biol.* **3**, 31 (1961); *ibid.* **4**, 161 (1962).
41. H. Kubinski, Z. Opara-Kubinska, W. Szybalski, *ibid.* **20**, 313 (1966).
42. K. A. Walsh and H. Neurath, *Proc. Nat. Acad. Sci. U.S.* **52**, 884 (1964); K. Brew, T. C. Vanaman, R. L. Hill. *J. Biol. Chem.* **242**, 3747 (1967); E. Freese and A. Yoshida, in *Evolving Genes and Proteins*, V. Bryson and H. Vogel, Eds. (Academic Press, New York, 1965), p. 341.
43. C. B. Bridges, *J. Heredity* **26**, 60 (1935).
44. G. P. Attardi, P. C. Huang, S. Kabat, *Proc. Nat. Acad. Sci. U.S.* **53**, 1490 (1965); *ibid.* **54**, 185 (1965); F. Ritossa, K. Atwood, D. Lindsley, S. Spiegelman, *Nat. Cancer Inst. Monogr.* **23**, 449 (1966); S. A. Yankofsky and S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.* **48**, 1466 (1962); H. Wallace and M. Birnstiel, *Biochim. Biophys. Acta* **114**, 296 (1966).
45. P. Borst, G. Ruttenberg, A. M. Kroon, *Biochim. Biophys. Acta* **149**, 140, 156 (1967); I. B. Dawid and D. R. Wolstenholme, *Biophys. J.*, in press; S. Granick and A. Gibor, in *Progr. Nucleic Acid Res. Mol. Biol.* **6**, 143 (1967).
46. C. A. Thomas, in *Progr. Nucleic Acid Res. Mol. Biol.* **5**, 315 (1966).
47. R. B. Church and B. J. McCarthy, *J. Mol. Biol.* **23**, 459, 477 (1967); D. D. Brown and J. B. Gurdon, *ibid.* **19**, 399 (1966); H. Wallace and M. Birnstiel, *Biochim. Biophys. Acta* **114**, 296 (1966); A. H. Whiteley, B. J. McCarthy, H. R. Whiteley, *Proc. Nat. Acad. Sci. U.S.* **55**, 519 (1966); H. Denis, *J. Mol. Biol.* **22**, 269, 285 (1966); V. R. Glisin, M. V. Glisin, P. Doty, *Proc. Nat. Acad. Sci. U.S.* **56**, 285 (1966); M. Nemer and A. A. Infante, *Science* **150**, 217 (1965); A. Spirin and M. Nemer, *ibid.*, p. 214; M. Crippa, E. Davidson, A. E. Mirsky, *Proc. Nat. Acad. Sci. U.S.* **57**, 885 (1967).
48. W. Gray, W. Dreyer, L. Hood, *Science* **155**, 465 (1967); S. Cohen and C. Milstein, *Nature* **214**, 449 (1967); G. M. Edelman and J. A. Gally, *Proc. Nat. Acad. Sci. U.S.* **57**, 356 (1967).
49. F. Crick (personal communication) has proposed that the repetitive segments of the DNA play their role structurally through interaction of these particular nucleotide sequences with the proteins of the chromosome.
50. G. G. Simpson, *This View of Life* (Harcourt, Brace and World, New York, 1964), p. 164.

51. We thank M. Chamberlin for excellent technical assistance; our colleagues R. B. Roberts, D. B. Cowie, E. T. Bolton, D. J. Brenner, and A. Rake for direct participation in the preparation of this article; B. H. Hoyer, D. Axelrod, and M. Martin for helpful discussion and for several DNA preparations; R. B. Roberts for proposing the schematic diagram (Fig. 13); F. R. Boyd and J. L. England for assistance in the design of the high pressure pump; and our colleagues and wives for patience and forbearance.