Characterization of Families of Repeated DNA Sequences from Four Vascular Plants†

Arnold J. Bendich* and Robert S. Anderson

ABSTRACT: Reassociation kinetics were measured spectrophotometrically at several temperatures for DNA from barley, daffodil, deer fern, and parsley fern. The data indicate that several to many families of related DNA base sequences in a single kinetic component comprise about half of the genome in these plants. The various families are not related to one another. They are predominantly *homogeneous* because each family contains member sequences related by the same degree of similarity. In one family the members would all be related by, for example, 95% sequence homology. In order to facilitate the interpretation of data for plant DNA, we investigated the parameters of DNA reassociation with simple DNAs. When bacterial or bacteriophage DNA was cooled from high temperature to the temperature for reassociation, a rapid decrease in absorbance termed "collapse" hypochromicity was observed. We found that collapse depends upon the temperature of incubation, salt concentration, and base composition of DNA and have derived an equation for computing the collapse once the melting temperature of DNA is known. Reassociation kinetics for randomly sheared bacterial and bacteriophage DNAs were found to deviate from ideal second-order form when about half the DNA was reassociated. The form of the experimental curves for these simple DNAs was taken to represent the course of reassociation of randomly sheared DNA with no detectable repeated sequences and was used to assess the fraction of plant DNA that reassociated as a single kinetic component. The evolutionary history of repeated sequences in plants is discussed as is the question of whether sequences designated by investigators as "repeated" are, in fact, accepted by the cell as repetitious in some functional sense. We conclude that most repeated sequences may not be representative of functionally repetitious DNA and raise the possibility that such DNA may be useful to the cell as sequence-independent "filler".

The nuclei of eukaryotic organisms contain DNA base sequences represented many times. Since this feature of the genome has not been observed in prokaryotes (Britten and Kohne, 1968; Kato et al., 1974; our unpublished results), the so-called "repeated" DNA sequences would appear important to the eukaryotic way of life. A group of sequences similar enough to form a stable duplex (reassociate) under specified conditions of measurement is said to constitute a "family" of sequences (Britten and Kohne, 1968).

The number of sequences classified as members of families is not an intrinsic property of the DNA, but may vary with the conditions of measurement, since only rarely do such sequences approach complete base sequences repetition. Perhaps it is because this has not been sufficiently appreciated that little has been reported concerning the relationships among the nominally "repeated" sequences. Two types of families of sequences may be considered. A *heterogeneous family* (Figure 1a) contains member sequences of varying similarity ranging from nearly perfect replicas to sequences barely similar enough to reassociate with one another at the criterion of stringency set by temperature and salt. As the stringency of reassociation measurement is raised (by increasing temperature, for example), the size of heterogeneous families decreases since the distantly related members are no longer sufficiently similar to interact. The rate of reassociation depends on the concentration of interacting sequences and will decrease with increasing temperature (Figure 1b). A *homogeneous family* (Figure 1c) contains member sequences of the same similarity. In one homogeneous family the members are all related by, for example, 80% sequence homology; in a second by 85%, and so on. For a group of homogeneous families, as the stringency (temperature) is raised, entire families are operationally removed from the "repeated" sequence class. The size of the remaining homogeneous families is unaltered, as is their rate of reassociation (Figure 1d). We report on reassociation kinetics measured at several temperatures for DNA from four vascular plants. We conclude from the data that families of sequences in these plants are predominantly of the homogeneous type.

Materials and Methods

Preparation of DNA. Young fronds of parsley fern [*Cryptogramma crispa* (L.) R. Br.] and deer fern [*Blechnum spicant* (L.) Roth] were collected near Seattle, Wash. Daffodil (*Narcissus pseudonarcissus* L., cv. King Alfred) bulbs and barley (*Hordeum vulgare* L., cv. Trebi) seeds were germinated in moist Pearlite and green shoots were harvested when they were 5–10 cm long. Tissue was stored at −20 °C and DNA was extracted as described earlier (Bendich and Anderson, 1974; Bendich and Bolton, 1967) with the addition of 0.1 M sodium diethyldithiocarbamate to the extraction buffer for fern tissue. This agent prevented the fern extracts from turning dark brown. Labeled barley DNA was extracted without prior freezing from roots plus shoots of 20 seedlings grown in the absence of bacteria (Bendich and Anderson, 1974) for 2 days in 1 mCi of [^{3}H]thymidine (59 Ci/mmol; New England Nuclear) with minimal irrigation.

DNA was extracted from *Escherichia coli* B, *Bacillus subtilus* 746, *Bacillus megaterium*, and *Pseudomonas aeruginosa* by Marmur's (1961) method. DNA from the *Agrobacterium tumefaciens* phage PS8 was the gift of T. Currier. T4 phage was the gift of J. Levy, and its DNA was extracted with phenol.

† From the Departments of Botany and Genetics, University of Washington, Seattle, Washington 98195. *Received May 18, 1977.* This work was supported by National Science Foundation Grant No. GB41179 and National Institutes of Health Grant No. GM22870-01.

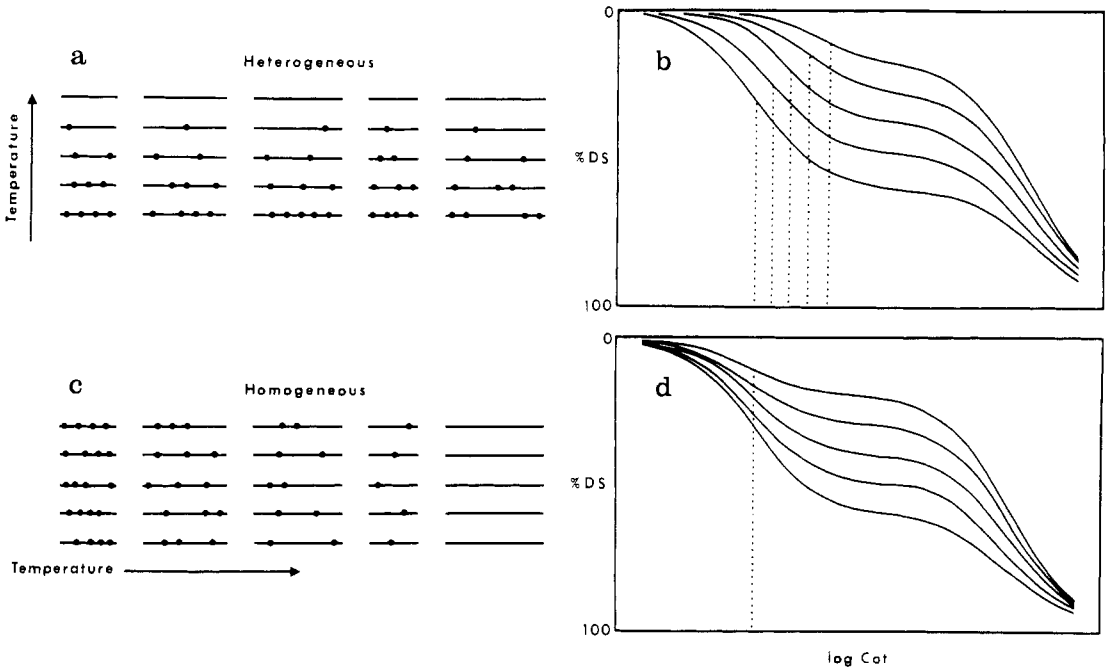

FIGURE 1: The difference between heterogeneous and homogeneous families of repeated DNA sequences. Each sequence represented by a line is a member of a vertically arrayed family. Members of one family are sufficiently similar to reassociate with each other but not with members of other families. Dots on the lines represent base changes. As the temperature of reassociation is increased, the more divergent sequences can no longer reassociate and the fraction of rapidly reassociating DNA decreases. For heterogeneous families (a) increasing temperature decreases the size of each family which decreases the rate of reassociation of the fast (repeated sequence) component. Thus, as the size of the early component decreases with increasing temperature, its $C_0t_{1/2}$ increases (b). For homogeneous families (c) increasing temperature eliminates entire families but the size and rate of reassociation of the remaining families are unaltered. Thus, as the size of the early component decreases with increasing temperature, its $C_0t_{1/2}$ remains constant (d). Rate effects due to base mispairing and suboptimal temperature have been disregarded for the sake of simplicity.

DNA was sheared to an average length of about 1100 nucleotide pairs (np),[1] as determined by electron microscopy and agarose gel electrophoresis, with a French pressure cell and to about 350 np (determined by gel electrophoresis) by sonication.

Melting and Reassociation. Measurements were made in Gilford recording spectrophotometers equipped with the thermal programmer and reference compensator accessories. Dialyzed samples were heated at 1 °C/min (0.25 °C/min for tetraethylammonium chloride solvent) to maximum hyperchromicity before cooling to the incubation temperature for reassociation. This cooling required about 2.0, 1.7, 1.5, 1.3, 1.0, 0.7, and 0.4 min to reach about 35, 30, 25, 20, 15, 10, and 5 °C, respectively, below the melting temperature (t_m) of native DNA. Rate and Δt_m (difference in t_m between native and reassociated DNA) data for plant DNA have been standardized against *Bacillus subtilis* DNA analyzed simultaneously in each case to account for temperature effects on rate and for thermal damage. Rate data have been corrected for the effect of base mispairing (indicated by the Δt_m) as recommended by Marsh and McCarthy (1974). DNA concentrations for rate measurements were determined on the assumption that the hyperchromic effect of native DNA is 38%.

Hydroxylapatite Fractionation. Hydroxylapatite powder (Bio-Gel hydroxylapatite from Bio-Rad) was mixed with an equal weight of cellulose powder (Whatman CF11) for increased flow rates and suspended in 10 mL of buffer per g of mixed powders. One milliliter of this slurry was used to pack each column in a Pasteur pipet. Sonicated ^{3}H-labeled barley DNA was stripped of foldback DNA by denaturation at 29 μg/mL in 1 mM NaCl–5 mM Tris buffer (pH 8.0) followed by rapid cooling in an ice bath. The solution was then brought to 1 mL in 120 mM potassium phosphate (pH 6.8) (KP) at 5.5 μg/mL and passed over a hydroxylapatite column at 50.6 °C. The column was washed with 1.5 mL of 120 mM KP and denatured DNA was recovered as unbound material. This temperature and salt condition represents 35 °C below the t_m (designated as $t_m - 35$ °C). It required 3–4 min for the solution to pass through the column, yielding a C_0t (Britten and Kohne, 1968) of 4×10^{-3}, or an equivalent C_0t in 1 M $NaClO_4$ of about 4×10^{-4}.

Foldback-stripped ^{3}H-labeled barley DNA at 180 μg/mL was denatured at 103 °C for 3 min, reassociated at 53.5 °C ($t_m - 35$ °C) to $C_0t = 1.6$ in 1 M $NaClO_4$–30 mM Tris buffer (pH 8.0), and loaded (with a 1.5-mL wash) onto hydroxylapatite in 120 mM KP at 50.6 °C. The 63% of the ^{3}H which bound to the column was eluted with 1.5-mL washes of 80 mM KP at each indicated temperature in order to construct a thermal elution (melting) profile of the reassociated DNA. To the indicated melting fractions was added 7 μg of denatured *B. subtilis* DNA as carrier and the fractions were dialyzed to 1 mM Tris buffer, dried in an air stream, redissolved in 1 M $NaClO_4$/Tris, and reassociated at 2.4 μg/mL at 53.5 °C to $C_0t = 0.5$ (or a C_0t of 0.5/0.63 = 0.8, if calculated on a total DNA basis). Radioactivity was measured in a Packard scintillation counter by mixing 5 mL of scintillant (1 vol of Triton X-100 plus 2 vol of Omnifluor dissolved in toluene) per mL of aqueous sample.

Results

The Starting Point of Reassociation: "Collapse" Hypochromicity. The reassociation of DNA strands was followed

[1] Abbreviations used are: np, nucleotide pairs; t_m, temperature at which half the DNA has been thermally denatured (the melting point of DNA); $t_m - n°$, n degrees below the t_m; Δt_m, difference in t_m between native and reassociated DNA; KP, potassium phosphate buffer, pH 6.8; Tris, tris(hydroxymethyl)aminoethane; C_0t, product of molar concentration of DNA nucleotides and time of incubation [(mol s)/L]; Et_4NCl, tetraethylammonium chloride.

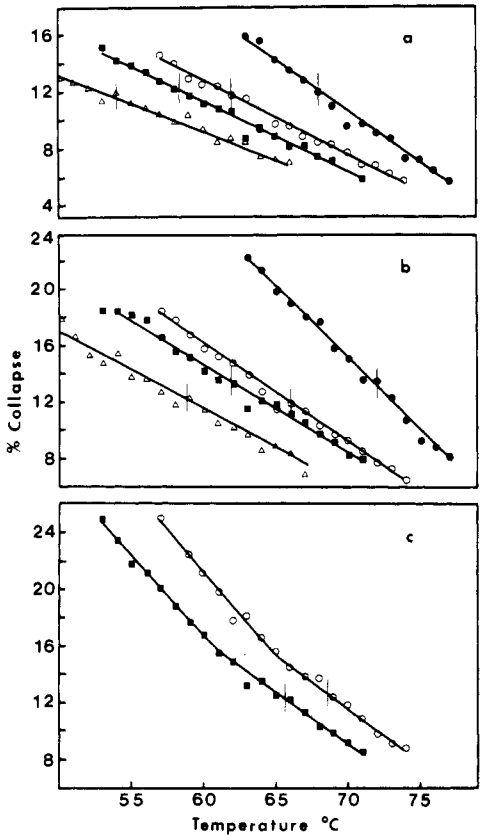

FIGURE 2: Collapse hypochromicity as a function of temperature in 0.07 M sodium phosphate (a), 0.14 M sodium phosphate (b), and 1 M sodium perchlorate (c). Collapse is expressed as the percent of the hyperchromic effect of native DNA upon melting. The source of DNA, average length 1100 np, was T4 bacteriophage (△), *B. subtilis* (■), *E. coli* (○), and *Pseudomonas aeruginosa* (●). The vertical lines indicate the collapse at $t_m - 25$ °C.

in the spectrophotometer by hypochromicity measurement. It was therefore necessary to distinguish between absorbance loss due to bimolecular reassociation from that caused by increased secondary structure when denatured DNA is cooled from about 100 °C to the temperature of incubation. This latter "collapse" hypochromicity was determined at a series of temperatures using bacterial and phage DNAs.

DNA at low concentration was melted and quickly brought to a given temperature while absorbance was monitored. After 2-3 min, the transition between the rapid (collapse) and the gradual loss of absorbance due to bimolecular reassociation was clearly evident. The DNA was again melted and the cycle was repeated for various temperatures. Each point in Figure 2 represents one such cycle where collapse is expressed as the percent of the increase in the absorbance of native DNA upon melting (the hyperchromic effect). For low (Figure 2a) and moderate (Figure 2b) salt concentration the collapse is a linear function of temperature in the temperature range investigated. The slopes of the lines increase with increasing base composition from 33 to 65% G + C, but at 25 °C below the melting point of native DNA (designated as $t_m - 25$ °C; indicated by a vertical line) the collapse is about 12%. For high salt concentration (Figure 2c) the curves appear biphasic, but again at $t_m - 25$ °C the collapse is about 12%.

Figure 3 presents the slopes of the lines of Figure 2 as a function of base composition. For 1 M $NaClO_4$ only the slope of the linear component at high temperatures (down to about $t_m - 30$ °C) was used in Figure 3. The following equation

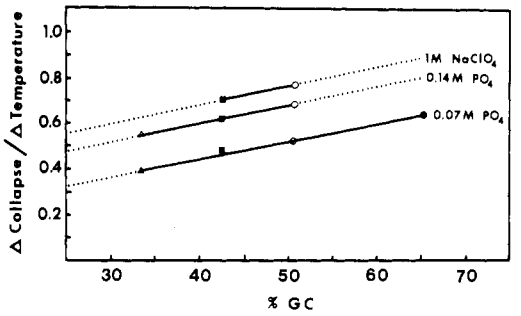

FIGURE 3: The change in collapse hypochromicity with temperature as a function of base composition. The slopes of the lines in Figure 2 are plotted against the base composition of DNA. Base composition was calculated from the t_m relative to that of *E. coli* DNA taken as 50.5% G + C. For 1 M $NaClO_4$ the slopes were taken from the linear components at high temperatures. The symbols used are as in Figure 2.

describes the relationship between collapse and base composition for the three salt concentrations:

$$\% \text{ collapse} = a + (t_{25} - t_i)(0.0078GC + b)$$

where a is the percent collapse at $t_m - 25$ °C and equals 11.9, 12.6, and 12.4 for 0.07 M NaP, 0.14 M NaP, and 1 M $NaClO_4$, respectively; t_{25} is the temperature at $t_m - 25$ °C; t_i is the incubation temperature used for reassociation; GC is the percent G + C determined from the melting point; b is the ordinate intercept at 0% G + C and equals 0.13, 0.31, and 0.38 for the three salts, respectively. For 1 M $NaClO_4$ at lower temperatures (from about $t_m - 30$ °C to $t_m - 37$ °C) a equals 12.4, b equals 0.46, and $0.0078GC$ is changed to $0.0143GC$.

It can be seen that the collapse depends upon salt, temperature, and base composition, but one need only determine the melting point to calculate the expected collapse using the equation. The data from which the equation was derived span the range between 12 and 37 °C below the t_m of native DNA.

Reassociation Kinetics for Simple DNA. Reassociation should take the form of a second-order reaction if all the DNA fragments terminate at the same points in the sequence, have the same length, and contain no internal repetitions (Britten and Davidson, 1976a). However, the reassociation kinetics of randomly sheared DNA deviate from second order since duplex structures almost always contain single-stranded regions which will subsequently interact with single strands and single-stranded regions at a reduced rate (Smith et al., 1975; Britten and Davidson, 1976a). We therefore investigated the form of reassociation kinetics with bacterial and phage DNAs so as to facilitate the interpretation of plant DNA reassociation.

Reassociation kinetics at $t_m - 25$ °C of randomly sheared DNA were monitored optically and the data are presented in Figure 4. The Scatchard plot is used since it facilitates determining the termination of the reaction (Marsh and McCarthy, 1973). Figure 4 also shows the calculated curve of an ideal second-order reaction. The experimental curves fail to go to complete reassociation presumably due to severe retardation of rate late in the reaction as polymeric structures form.

The data for the four experimental curves in Figure 4 were replotted as C_0t curves (Britten and Kohne, 1968) along with a calculated ideal second-order curve. The deviations of each of the four from the ideal curve were then used to construct a curve reflecting the average deviation. This constructed curve is depicted in Figure 5 along with an ideal curve. The constructed curve approaches 90% and begins to deviate from the ideal curve at about 50% reassociation. Gillis and De Ley

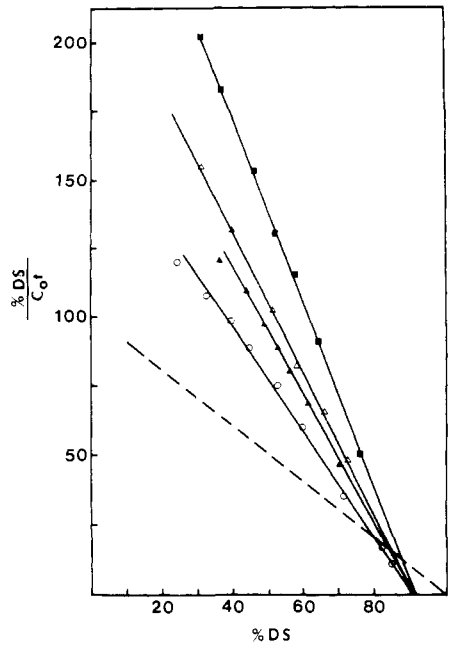

FIGURE 4: Reassociation kinetics for simple DNA. Optical measurements were made at $t_m - 25$ °C using 1100 np fragments from T4 phage (■, 25 μg/mL in 0.14 M sodium phosphate), PS8 phage (○, 13 μg/mL in 0.14 M sodium phosphate), *B. subtilis* (△, 69 μg/mL in 1 M $NaClO_4$–30 mM Tris, pH 8), and *B. megaterium* (▲, 87 μg/mL in 0.15 M NaCl–0.015 M sodium citrate, pH 7). An ideal second-order curve intercepts the abscissa at 100% as is shown by the broken line. The extrapolation curves for the experimental data intercept the abscissa at 90 to 91%. The term % DS represents the percent of DNA in double-stranded form.

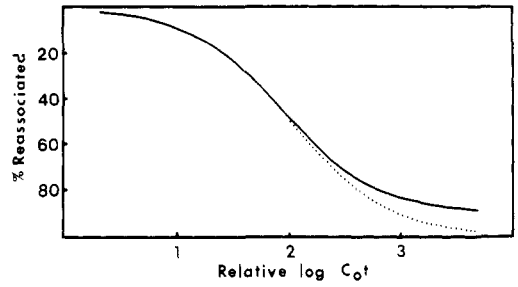

FIGURE 5: Departure of reassociation rate data from ideal second-order kinetics. The solid line shows the average deviation of the experimental data in Figure 4 from an ideal second-order curve (the broken line).

(1975) have also noted the departure of experimental data from second-order kinetics at about 50% reassociation. We have taken the constructed curve in Figure 5 to represent the course of reassociation of a randomly sheared DNA containing a single kinetic component: a "pseudoideal" curve. We call such a curve a 100% component and have calculated a series of standard curves at 5% intervals for use in determining the percent of a C_0t curve which is describable as a single kinetic component. Thus, if data points from a plant DNA C_0t curve fall on a 50% component curve, we assume that half the plant DNA reassociates with kinetics as closely approximating second order as do bacterial and phage DNAs. The half-point for reassociation, the $C_0t_{1/2}$, of a 100% component curve (solid line in Figure 5) is taken at 50% (not 45%) on the ordinate since the kinetics of all the DNA is to be assessed regardless of rate reduction late in the reaction.

We chose to use short fragment lengths for studying plant DNA reassociation in order to avoid possible difficulties arising from networks or aggregates which form late in the reaction using moderate to long fragments (Thompson, 1976; our unpublished results). Using *B. subtilis* DNA fragmented by sonication to about 350 np the extrapolated completion of seven reassociation kinetics measurements at $t_m - 25$ °C averaged 92% with a range of 90–95%, in agreement with the data in Figure 4 for 1100 np fragments. When measurements were made at $t_m - 10$ °C and $t_m - 5$ °C, the extrapolated completion was about 80 and 60%, respectively. Consequently we find that the component curves which best fit *B. subtilis* data are 100, 90, and 65% for the criteria of $t_m - 25$ °C, $t_m - 10$ °C, and $t_m - 5$ °C, respectively.

Reassociation Kinetics for Plant DNA. A survey was made of 20 vascular plants including flowering plants, a gymnosperm, ferns, psilophytes, and *Equisetum* (horsetail) species by measuring reassociation kinetics at $t_m - 25$ °C. All of these species contained rapidly reassociating DNA sequences typical of higher eukaryotes. Four species which exhibited a single repeat frequency component comprising about half the total DNA were selected for further analysis. In all measurements *B. subtilis* DNA was included as one of the three DNAs analyzed simultaneously in the spectrophotometer. This bacterial DNA had a t_m within 1 °C of that for each plant DNA and served as a standard to account for temperature effects on rate. For example, if the reassociation of *B. subtilis* DNA at $t_m - 5$ °C followed a 65% component curve with a $C_0t_{1/2}$ 1.4-fold that at $t_m - 25$ °C, the $C_0t_{1/2}$ values of the plant DNAs analyzed simultaneously were corrected by the factor necessary to adjust the bacterial curve to its $C_0t_{1/2}$ observed at $t_m - 25$ °C (1/1.4). Collapse hypochromicity was calculated as described above.

The reassociation kinetics at $t_m - 25$ °C of DNA from the four plants are depicted as C_0t curves in Figure 6. Data points from measurements using DNA concentrations varying over a 30- to 80-fold range fall on the same curves indicating the reactions are bimolecular (Wetmur, 1976). We were unable to detect any very rapidly reassociating components in the unfractionated DNAs used here. Given the sensitivity of our measurements we conclude that there is less than about 2–3% of total DNA which reassociates extremely rapidly. The curves drawn through the data points at low C_0t values are pseudoideal curves (see the previous section) for 55, 50, 45, and 45% components for parsley fern, barley, daffodil, and deer fern, respectively. Each DNA also appears to have a second frequency component which is fit by a pseudoideal curve for 20, 25, 30, and 30% components for these plants, respectively. The remaining 25% of each genome reassociates slowly at $t_m - 25$ °C. We did not attempt to continue the reassociation in order to assess the rate for the slow components.

A series of reassociation rate measurements was then made for each of the four plant DNAs at several temperatures ranging from about $t_m - 35$ °C to about $t_m - 5$ °C. *B. subtilis* DNA was analyzed simultaneously in each case for standardization. One series of C_0t curves for each plant is shown in Figure 7. Pseudoideal curves to the nearest 5% are fitted to the data points which are arbitrarily displaced laterally in order to facilitate inspection. The pseudoideal curves appear to closely approximate the data points indicating the presence of a single frequency component under these conditions of stringency. The results of all such measurements are presented in Figure 8. The fraction of total DNA which appears in the fast component decreases gradually from about $t_m - 35$ °C to about $t_m - 20$ °C or $t_m - 15$ °C and then decreases more precipitously as the temperature of reassociation approaches the t_m. However, the $C_0t_{1/2}$ of the fast component does not change much over this temperature range. If the families of repeated sequences in these plants were of the heterogeneous type (Figure 1), the $C_0t_{1/2}$ would be expected to increase in

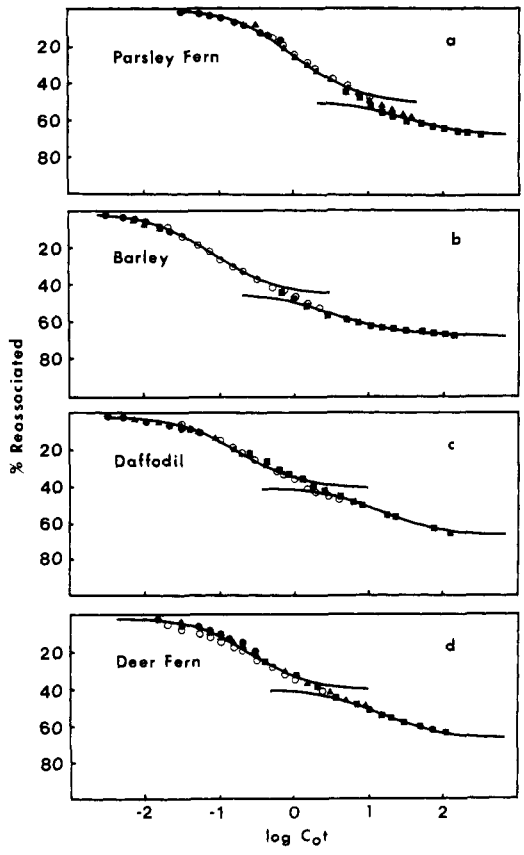

FIGURE 6: Reassociation kinetics of plant DNAs at $t_m - 25$ °C. Optical measurements were made in 1 M $NaClO_4$–30 mM Tris (pH 8) using 350 np fragments: (a) parsley fern at 10 (●), 78 (○)*, 80 (▲), and 870 (■) μg/mL; lines are pseudoideal curves for 55 and 20% components; (b) barley at 10 (●), 22 (▲), 73 (○)*, and 840 (■) μg/mL; lines are 50 and 25% components; (c) daffodil at 12 (●), 21 (▲), 78 (○)*, and 610 (■) μg/mL; lines are 45 and 30% components; (d) deer fern at 36 (●), 67 (○)*, 85 (▲), and 1000 (■) μg/mL; lines are 45 and 30% components. The data indicated by the asterisk (*) are those used for the $t_m - 25$ °C points in Figure 8.

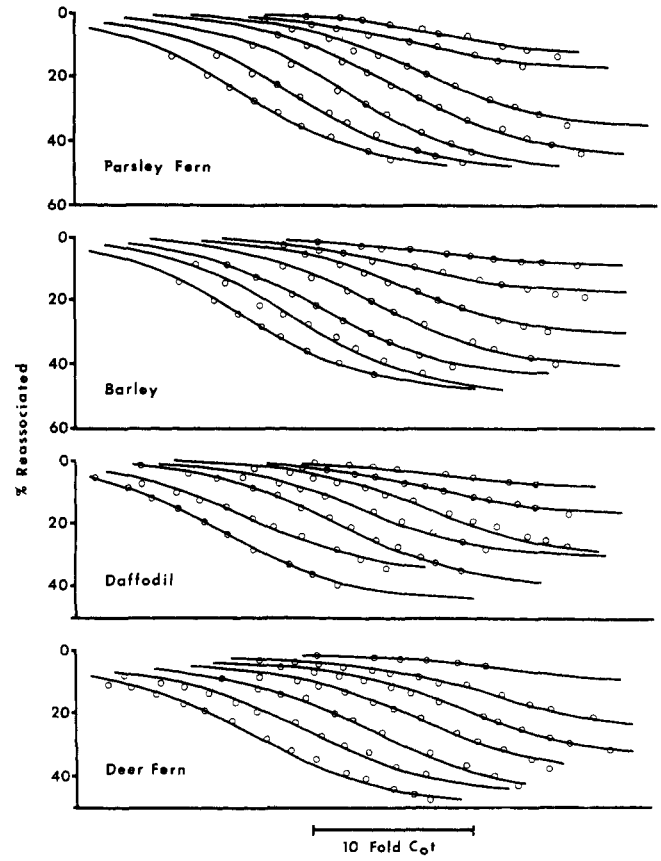

FIGURE 7: Reassociation kinetics of plant DNAs at several temperatures. Optical measurements were made at 5 °C intervals between about $t_m - 35$ °C and about $t_m - 5$ °C in 1 M $NaClO_4$–30 mM Tris (pH 8) with 350 np fragments at 67 to 90 μg/mL. Lines are pseudoideal curves for components (to the nearest 5%) best fitting the data points which are arbitrarily displaced laterally. Collapse hypochromicity for measurements at $t_m - 5$ °C and $t_m - 10$ °C was readily evident from the abrupt transition between the rapid and gradual loss of absorbance upon cooling of *B. subtilis* DNA to the incubation temperature. The equation given in the text was used to calculate collapse at the other temperatures.

proportion to the decrease in the size of the fast component as the temperature approaches the t_m. From Figure 8 the ratio of fast component size at $t_m - 25$ °C to that at $t_m - 5$ °C is 3.6, 4.4, 4.4, and 4.1 for parsley fern, barley, daffodil, and deer fern, respectively. Yet the increases in $C_0t_{1/2}$ over this temperature range are only 1.6-, 1.8-, 2.4-, and 1.3-fold for these plants, respectively. These results indicate that the families of repeated sequences in the four plants are predominantly of the *homogeneous* type. Included in Figure 8 are data from rate measurements employing 2.4 M tetraethylammonium chloride (Et_4NCl). Melchior and Von Hippel (1973) and Chang et al. (1974) have shown that the t_m of DNA is independent of base composition in this solvent. In a study of phage DNA reassociation, Chang et al. (1974) concluded that this is the preferred solvent for measuring DNA reassociation. In an analogous series of measurements we used DNA from the same four plants, but at a fragment length of about 1100 np (Bendich, 1977). Fewer measurements were made and only the 1 M sodium perchlorate solvent was used in this earlier work. Nevertheless the results were similar to the present results in that no increase in $C_0t_{1/2}$ was observed with increasing temperature of reassociation.

The Δt_m provides a quantitative estimate of the degree of mispairing present in reassociated DNA strands (Ullman and McCarthy, 1973). The Δt_m for each DNA reassociated at the various temperatures is shown in Figure 9 as a function of the temperature of reassociation. The shapes of the curves are similar for the four plants and show no discontinuities. Thus, for each plant there appears to be a smooth gradation of homogeneous families ranging from those with nearly perfectly matched members (Δt_m about 1 °C) to families composed of distantly related members with a large Δt_m. Wetmur (1976) has reviewed the relationship between Δt_m and percent mispairing and has concluded that each percent mispairing of bases causes an increase in Δt_m of about 1.1 °C.

A homogeneous family of the type depicted in Figure 1 which exhibits a high Δt_m does not contain member sequences which will reassociate to yield a low Δt_m. A homogeneous family containing closely related sequences will reassociate to form only duplexes of high thermal stability. On the other hand, a heterogeneous family contains sequences which will reassociate with both closely and distantly related family members. To test whether closely and distantly related sequences present at the same reiteration frequency can cross-react, the following experiment was performed. Tritium-labeled barley DNA was stripped of foldback sequences, denatured, and incubated at $t_m - 35$ °C to a C_0t of 1.6, a point at which essentially all of the fast component would have reassociated (see Figure 6b). The resulting duplexes were bound to and thermally eluted (melted) from hydroxylapatite (Figure 10). Fractions eluting at five temperatures were again reassociated at $t_m - 35$ °C and remelted from hydroxylapatite. It

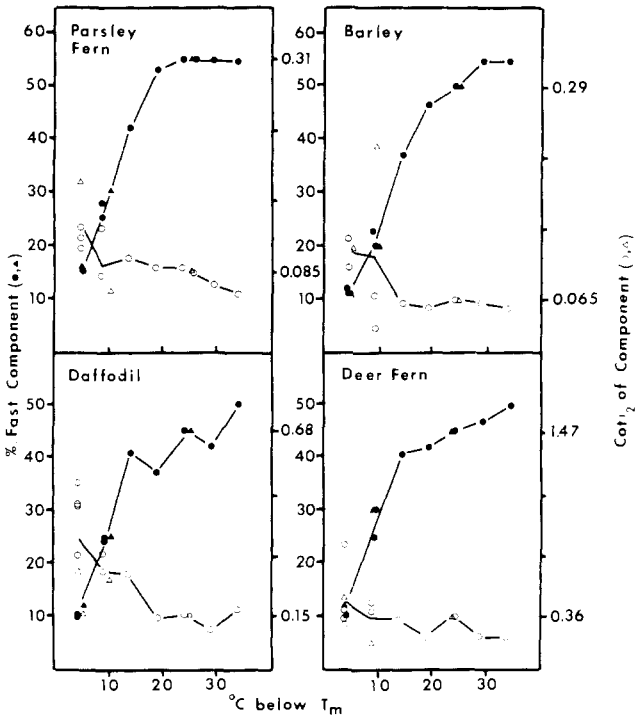

FIGURE 8: Summary of reassociation data for plant DNAs. From the data in Figure 7 and analogous measurements the size and the $C_0t_{1/2}$ of the fast component are plotted against degrees below the t_m for reassociation. The data have been adjusted for temperature effects on rate by using $t_m - 25$ °C as a standard and for the effect of base mispairing (Δt_m) on rate. We assumed that the reassociation of *B. subtilis* DNA followed 65, 90, and 100% component curves for measurements at $t_m - 5$ °C, $t_m - 10$ °C, and $t_m - 15$ to 35 °C, respectively (see text). These were the "expected" component sizes. In some cases rate data at several temperatures were obtained by serially melting and reassociating the same DNA sample and therefore the data were adjusted for possible effects of thermal damage. The triangles are data obtained in 2.4 M Et_4NCl; circles are data for 1 M $NaClO_4$–30 mM Tris (pH 8). Collapse in Et_4NCl at $t_m - 25$ °C was taken as 12.4%. The $C_0t_{1/2}$ for *B. subtilis* at $t_m - 25$ °C was 2.1 and 1.9 in Et_4NCl and 0.81, 0.68, 0.78, 0.75, and 0.86 (mean 0.78) in $NaClO_4$. The $C_0t_{1/2}$ data for Et_4NCl have been adjusted for this difference in rate. The values given on one ordinate are the $C_0t_{1/2}$ at $t_m - 25$ °C and this value multiplied by the ratio of the component size at $t_m - 25$ °C to that at $t_m - 5$ °C. The greater value thus indicates the $C_0t_{1/2}$ expected at $t_m - 5$ °C if the families were heterogeneous. Calculations and data adjustments were made as follows, using *B. subtilis* DNA analyzed simultaneously as a reference. (1) The ratio of the observed to the expected component size for *B. subtilis* was computed. This value, termed A, was 1.0 except when serial analyses were made with the same DNA sample; it then was 0.77, 0.83, or greater. (2) The observed $C_0t_{1/2}$ for *B. subtilis* was multiplied by A. This compensates for the expectation that thermally damaged DNA ("tired DNA") contributes to C_0 by absorbing, but does not reassociate completely. (3) The adjusted $C_0t_{1/2}$ for *B. subtilis* DNA at $t_m - 25$ °C was assigned the value 1.0 and the *B. subtilis* data at other temperatures were then normalized to $t_m - 25$ °C. The term B refers to the normalization factor for each temperature and was less than 1.8. (4) The observed component size for plant DNA was divided by A and the resulting values are given in the figure. (5) The observed $C_0t_{1/2}$ for plant DNA was multiplied by A to compensate for "tired DNA". (6) The resulting $C_0t_{1/2}$ was corrected for the effect of base mispairing as recommended by Marsh and McCarthy (1974) using the Δt_m values given in Figure 9. This correction changed the $C_0t_{1/2}$ at about $t_m - 5$ °C by 13% or less. (7) The $C_0t_{1/2}$ for plant DNA was finally multiplied by B and the resulting values are given in the figure.

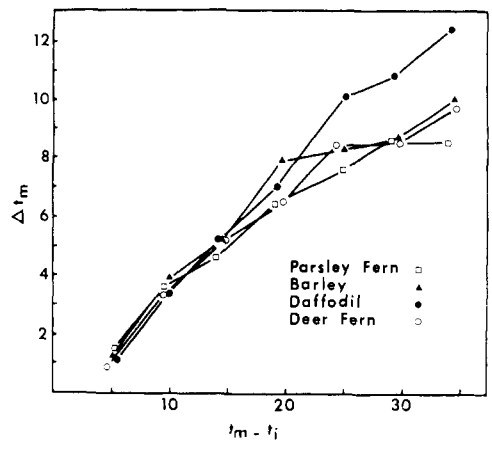

FIGURE 9: Change in thermal stability of DNA reassociated at various temperatures below the t_m of native DNA. The difference between the t_m of native and reassociated DNA (Δt_m) is plotted against the difference between the native t_m and the incubation temperature for reassociation ($t_m - t_i$). The Δt_m of *B. subtilis* DNA analyzed simultaneously (about 1 °C regardless of incubation temperature) has been subtracted from that for the plant DNAs to compensate for thermal damage so that the values in the figure more accurately reflect mispairing due to sequence divergence. The data are from the experiments also reported in Figure 8 and are mean values when more than one measurement was made.

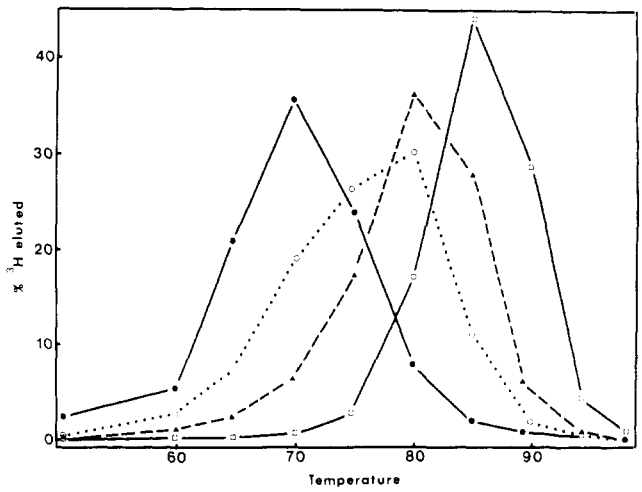

FIGURE 10: Thermal fractionation on hydroxylapatite of reassociated barley DNA. Tritium-labeled barley DNA (350 np, 20 000 cpm/μg) from which the foldback fraction (5% of the total) was removed at $t_m - 35$ °C was denatured and incubated at $t_m - 35$ °C to a C_0t of 1.6 and loaded onto hydroxylapatite at $t_m - 35$ °C. The bound DNA was thermally eluted in 0.08 M potassium phosphate (○). Fractions eluting at five temperatures were again reassociated at $t_m - 35$ °C to a C_0t of 0.5 (or a C_0t of 0.5/0.63 = 0.8, if calculated on a total DNA basis). A second melting profile was then constructed. For clarity only the remelts of the fractions eluting at the lowest (●) and highest (▲) of the five temperatures are shown, as is a profile for native DNA (□). The t_m values of the remelts were 67.8, 69.4, 72.5, 75.0, and 78.0 °C for the fractions eluting at 64.9, 70.1, 74.8, 80.0, and 85.0 °C, respectively. The t_m of the first melting profile was 73.7 °C, while that for native DNA was 83.2 °C.

can be seen that DNA which eluted at low temperature in the first melting profile reassociated to form predominantly low-melting duplexes. Reassociated DNA which eluted at high temperature in the first profile re-formed duplexes that were predominantly high melting despite the fact that both incubations were at the relatively low stringency criterion of $t_m - 35$ °C. Thus, well-matched and less-well-matched sets of repetitive DNA within a single frequency class show little cross-reaction. These results do not conform to the heterogeneous family model, but are consistent with the homogeneous family alternative. This experiment was not attempted with the other three plants.

Using absorbance measurement to assess reassociation we found in Figure 6b that less than 2% of the nucleotides in barley DNA were base paired by $C_0t = 3 \times 10^{-3}$. Yet the foldback fraction (C_0t less than 4×10^{-4}) contained 5% of the ^3H-labeled barley DNA initially 350 np in length. Thus, the

strands in the foldback fraction would contain less than half of their nucleotides in base-paired form.

Discussion

The major conclusion to be drawn from the data presented here is that most of the DNA in parsley fern, barley, daffodil, and deer fern that reassociates rapidly and is therefore classified as repetitious in a chemical sense exists in *homogeneous* families of sequences (see Figure 1). Reassociation rate measurements in the spectrophotometer provided most of the data leading to this conclusion. Thus, it was necessary first to investigate some of the parameters involved in the optical measurements and these will be discussed first.

Assessing the Beginning and Termination of Reassociation. The increase in secondary structure of DNA upon cooling from a temperature well above the t_m to the incubation temperature for reassociation is accompanied by a rapid loss of absorbance. This collapse hypochromicity was ignored by Wetmur and Davidson (1968) since it introduced only a 20% error in their calculations of the second-order rate constant for T4 phage DNA which has little or no detectable rapidly reassociating ("fast") components. However, when investigating the reassociation of complex DNAs by optical methods, it is important to account for collapse in order to assess the presence of fast components. Britten et al. (1974) estimated the collapse hypochromicity at about $t_m - 25$ °C to be 2.7% of the absorbance of denatured DNA at 97 °C. If we assume a 37% hyperchromic effect, this value becomes 10% of the hyperchromic effect which may be compared with our estimate of 12% for collapse at $t_m - 25$ °C. We are unaware of any previous systematic investigation of the effect of base composition of the DNA on collapse. And we know of no other work on collapse as a function of incubation temperature or of salt concentration. The equation derived above allows the collapse in three salt concentrations at many temperatures to be calculated when the t_m of DNA is known. It should be noted that even a small change in incubation temperature affects the collapse significantly. For example, in 0.14 M sodium phosphate buffer the collapse for a 40% G + C DNA at $t_m - 25$ °C is 12.6%, while at $t_m - 28$ °C the collapse is 14.5%. Unless this difference is taken into account, the greater loss of absorbance at the lower temperature could be mistakenly interpreted as indicating the presence of a small "fast component", when none is actually present, and as a result the assessment of the size and $C_0t_{1/2}$ for the true DNA component present will be inaccurate.

In this discussion we have considered collapse hypochromicity to reflect increased secondary structure at the lower temperature. However, two additional factors can also lead to a very rapid absorbance loss. Extremely fast components, such as some animal satellite DNAs, may be present which are completely reassociated before the temperature shift is complete. It is also possible that foldback sequences are present which contain sufficient base-pairing regions to contribute measurably to the rapid absorbance loss. The latter situation apparently does not exist for barley DNA. We have also assumed in this work that the absorbance loss due to increased base stacking at lower temperatures for bacterial and phage DNAs is equivalent to that for complex DNA, but have not tested this assumption experimentally.

We used the C_0t plot (Britten and Kohne, 1968) rather than the second-order plot (Wetmur and Davidson, 1968) for analyzing rate data for plant DNA because the former plot allows short-time and long-time data points to be placed on the same graph, whereas the latter does not. Since the plant genomes may contain several overlapping kinetic components, it is important to examine as much of the data as possible before making assignments of component size and $C_0t_{1/2}$ (the reciprocal of the rate constant).

We observed in our measurements with *B. subtilis* DNA at stringent criterion (about $t_m - 5$ °C) that the completion of the reassociation reaction, as estimated by extrapolation using the Scatchard plot, was about 60% rather than 90% which was observed at the less stringent criterion of $t_m - 25$ °C. This observation appears to indicate that either 40% of the nucleotide sequences in the bacterial DNA cannot reassociate at $t_m - 5$ °C or that 40% of the DNA reassociates only at a very much reduced rate at $t_m - 5$ °C. In the perchlorate solvent we noted that the hyperchromic effect in melting curves of native *B. subtilis* always began at a temperature below $t_m - 5$ °C. Therefore some of the DNA, probably in regions of low G + C content, should not be able to reassociate at $t_m - 5$ °C because this temperature exceeds the t_m of the putative low G + C regions. However, there seems to be a fragment length effect involved here as well. At $t_m - 5$ °C in 2.4 M Et_4NCl, a solvent in which t_m is independent of base composition, we still observed the extrapolated completion of reassociation to be 60% using 350 np fragments, but this value increased to 71 and 75% with fragment lengths of about 1100 and 5000 np, respectively; these values were at least 88% at $t_m - 25$ °C. The reassociation process with randomly sheared DNA fragments begins with a first collision and nucleation leading to a structure containing on the average only 55–60% of the length in duplex form and the remaining 45–40% of each strand unpaired (Smith et al., 1975; Wetmur, 1976). Subsequent reactions of single-stranded regions on duplex-containing structures become more important as reassociation continues but the length of available single strand will decrease. Since the t_m of DNA decreases in degrees by $500/L$ (Wilson and Thomas, 1973) or $650/L$ (Britten et al., 1974), where L is the duplex chain length in nucleotides, and since L could easily be reduced to 100 or less after the initial stages of reassociation with fragments averaging 350 np, it seems reasonable that the reaction should not go to completion at $t_m - 5$ °C with the 350 np fragments used in our study. It follows that unless the data are appropriately corrected (see Results section) to a standard at lower criterion, the kinetic complexity of a DNA obtained from measurements made at $t_m - 5$ °C will be overestimated and this effect will be most pronounced with short fragment lengths.

The Nature of Families of Repeated DNA Sequences and Their Evolutionary History. The results presented here do not rule out some heterogeneity among the families of sequences in the four plants under study. The $C_0t_{1/2}$ was found to increase somewhat for reassociation measurements conducted at temperatures approaching the t_m (Figure 8). However, the data do indicate that most of the sequences which reassociated at high temperature did not interact with more distantly related sequences at low temperature. If we assume, as did Britten and Kohne (1968), that families composed of divergent member sequences are older in an evolutionary sense than those of closely similar membership, one would expect some heterogeneity in the older families from statistical considerations. It is perhaps surprising that the older families are predominantly homogeneous. Apparently sequence divergence sufficient to produce an overall Δt_m of 8–12 °C did not produce much heterogeneity in these older families. Heterogeneity would increase if during the course of evolution some of the members of a given family were prevented from rapidly diverging while other members in the same family accumulated base changes at a more rapid rate. However it is likely, as Britten and Kohne (1968) have reasoned, that the nucleotide sequences of the members of the families are not conserved by severe selection. Rather than invoke a special selection

mechanism of this type for some of the thousands of members of any family, we proposed a more likely evolutionary history based on a constant rate of sequence change for all members of all families (Bendich, 1977). According to this scheme each family of sequences was created in a sudden or saltatory event and the genesis of a set of homogeneous families was an extended process involving many minor saltatory events over a long period of time. The families presently exhibiting a large Δt_m would then be older while those of small Δt_m would be very recently created. It was further proposed that a satellite or cryptic satellite DNA is the initial product of each such saltation and that subsequent translocation events scatter short lengths of repeated sequences throughout the genome. Evidence for homology between clustered repeats of satellite density and dispersed repeats of main-band density in the muskmelon has been presented (Bendich and Taylor, 1977; Bendich, 1977).

We suggest that the homogeneous families are the result of saltatory replications occurring at many different times with a constant rate of sequence change in all families. In order to account for the single kinetic component we observe, one of two requirements must be met. Either the parental sequence was copied the same number of times in each saltation or family members were discarded subsequent to their synthesis until a constant family size remained. It is not obvious why such a sequence amplification to a constant level should exist. Nor is it easy to formulate a mechanism by which the constancy is maintained. Nevertheless, a phenomenon termed "gene compensation" has been described which appears relevant to the question of constancy of sequence amplification. In *Drosophila* the number of genes for ribosomal RNA (Tartof, 1971) and for 5S RNA (Procunier and Tartof, 1975) is sometimes returned to a wild-type level after individual flies with various deletions for some of these genes are crossed. Thus, for sequences of defined function a mechanism apparently exists to control the degree of amplification. Perhaps it is not unreasonable to accept a controlled amplification of sequences whose function we do not understand.

We are unaware of any previous attempts to determine whether repeated sequence families in eukaryotic DNA are of the heterogeneous or homogeneous type (see, for example, the recent review by Britten and Davidson, 1976b). Data consistent with the homogeneous family alternative may be found in the work of Britten and Kohne (1968) in which salmon DNA was analyzed as in Figure 10 and showed little sequence homology between closely and distantly related sets of repetitive sequences. However these measurements could not themselves indicate the homogeneous family alternative because a single frequency class of sequences was not used in that work. Marsh (1974) monitored the reassociation of some animal DNAs in the spectrophotometer at several temperatures but did not use the data to assess the type of families present. In most published work on repeated sequences in eukaryotes only a single criterion for reassociation is employed, usually $t_m - 25$ °C. It is difficult to see how the question of whether families are heterogeneous or homogeneous could be approached without using several criteria for such measurements. It should be clear that a more complete understanding of the relationships among the related sequences can be achieved when measurements are made at several temperatures rather than at a single temperature.

On the Value of Sequence Repetition for the Investigator and for the Cell. In this section we will consider the recent direction of inquiry concerning sequence repetition, and whether it may lead us to an understanding of the biological role of repetitive DNA.

Let us assume for the moment that there are DNA sequences, other than those which specify the stable RNAs, that can function in concert because of their similarity and that such sequences are needed for orderly functioning of the organism. What degree of sequence homology would the cell require for such functional equivalence? Would the cell consider two stretches of DNA to be unrelated if they differ in sequence by 10%? By 1%? We have seen no discussion of this issue in the literature which deals with sequence repetition in eukaryotes. Investigators in this area usually set an operational definition for repetition by choosing $t_m - 25$ °C as the single criterion for their reassociation, although there have been some notable exceptions (Walker and McClaren, 1965; Martin and Hoyer, 1966; Bendich and McCarthy, 1970; McCarthy and Farquhar, 1972; Marsh, 1974). However, we do not know what level of discrimination is used by any eukaryote to distinguish repeat from nonrepeat. It would therefore be *fortuitous* if the properties of sequences classified as repetitious at $t_m - 25$ °C were to be representative of functionally repetitious DNA.

The data indicate that several to many homogeneous families comprise the single rapidly reassociating components in the plants studied here. Which of the families might be utilized for function? Since there is such a large proportion and amount of DNA in these families, it seems unlikely that all this DNA would function in a manner requiring sequence similarity. The families composed of more divergent members might be useful, but those with closely similar members should afford the advantage of more precise control over a postulated network of functions. We will assume for this argument that sequences must be nearly perfect replicas to be considered as reiterated by a cell. Just how similar the sequences need be is conjectural, but there are examples in prokaryotic systems in which a single base change can alter the recognition of a regulatory protein for its binding sequence (Gilbert et al., 1974; Maniatis et al., 1975; Dickson et al., 1977). There is, however, also an example in which rather divergent operator sequences are recognized by the λ repressor protein (Ptashne et al., 1976). In accord with our assumption, nearly all the families would be unsuitable for purposes which require functional repetition. Perhaps only some of the sequences which reassociated at $t_m - 5$ °C (0.1 °C Δt_m) would be sufficiently similar. Data for many organisms have been gathered from reassociation experiments conducted at $t_m - 25$ °C, the criterion of stringency adopted as standard by most investigators. These data have led to estimates of the length, spacing, number, and complexity of DNA sequences which appear to be repetitious at the standard criterion, as well as to assignments of the fraction of total DNA in such sequences. However, these properties and the generalizations derived from them may not be characteristic of functionally repetitious sequences. If functional repeats exist as a small number of nearly perfect replicas, and if their properties are different from those for the much larger amount of DNA which reassociates rapidly at the standard criterion, then the properties of functional repeats have been *obscured* by those of sequences regarded as reiterated by the investigator but not by the cell. The current of research may, in fact, be carrying us away from our goal of understanding the biological utility of repeated DNA sequences.

Repetitious DNA as Filler. Most authors in the area of repeated sequences seem to accept the criterion of $t_m - 25$ °C as reflective of functionally significant sequence repetition. Rarely do we find a differing point of view. But until we divine the *biological criterion* for sequence repetition we must either *assume* that this single experimental criterion is biologically representative, or we must entertain the possibility that sequences usually classified as repetitious may have no function

which requires similarity of sequence. That is, the cell does not utilize them as functionally equivalent and there are no metabolic events which are coordinated or regulated in a manner dependent on the sequence similarity so obvious to us. Viewed from this perspective the central issue may not be why we find so much nominally repetitious DNA, but why the genomes of higher organisms contain so much DNA. Others have speculated on why it may be advantageous for higher organisms to have DNA that functions in a manner independent of its sequence (Crick, 1971; Sutton, 1972; Lin and Riggs, 1975; Alberts et al., 1977). The advantage of a large genome, its generation by sequence amplification and dispersal, and the use of repetitious DNA in a sequence-independent "filler" capacity will be the subjects of a forthcoming publication.

Note Added in Proof

After this paper was accepted for publication, we learned of a paper [Bouchard, R. A., and Swift, H. (1977), *Chromosoma 61,* 317] in which it was concluded on the basis of experiments similar to those depicted in Figure 10 that the families of related sequences in the principal rapidly reassociating DNA component of the ferm *Thelypteris normalis* are of the *homogeneous* type, as defined in our paper.

References

Alberts, B., Worcel, A., and Weintraub, H. (1977), in Proceedings of the International Symposium on the Eukaryotic Genome, Bradbury, M., Ed. (in press).

Bendich, A. J. (1977), in Molecular Biology of Plants, Rubenstein, I., Ed., New York, N.Y., Academic Press (in press).

Bendich, A. J., and Anderson, R. S. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 1511.

Bendich, A. J., and Bolton, E. T. (1967), *Plant Physiol. 42,* 959.

Bendich, A. J., and McCarthy, B. J. (1970), *Genetics 65,* 545.

Bendich, A. J., and Taylor, W. C. (1977), *Plant Physiol. 59,* 604.

Britten, R. J., and Davidson, E. H. (1976a), *Proc. Natl. Acad. Sci. U.S.A. 73,* 415.

Britten, R. J., and Davidson, E. H. (1976b), *Fed. Proc., Fed. Am. Soc. Exp. Biol. 35,* 2151.

Britten, R. J., Graham, D. E., and Neufeld, B. R. (1974), *Methods Enzymol. 29E,* 363.

Britten, R. J., and Kohne, D. E. (1968), *Science 161,* 529.

Chang, C., Hain, T. C., Hutton, J. R., and Wetmur, J. G. (1974), *Biopolymers 13,* 1847.

Crick, F. (1971), *Nature (London) 234,* 25.

Dickson, R. C., Abelson, J., Johnson, P., Resnikoff, W. S., and Barnes, W. M. (1977), *J. Mol. Biol. 111,* 65.

Gilbert, W., Maizels, N., and Maxam, A. (1974), *Cold Spring Harbor Symp. Quant. Biol. 38,* 845.

Gillis, M., and De Ley, J. (1975), *J. Mol. Biol. 98,* 447.

Kato, A. C., Borstad, L., Fraser, M. J., and Denhardt, D. T. (1974), *Nucleic Acids Res. 1,* 1539.

Lin, S., and Riggs, A. D. (1975), *Cell 4,* 107.

Maniatis, T., Ptashne, M., Backman, K., Flashman, S., Jeffrey, A., and Mauer, R. (1975), *Cell 5,* 109.

Marmur, J. (1961), *J. Mol. Biol. 3,* 208.

Marsh, J. L. (1974), Ph.D. Thesis, University of Washington, Seattle.

Marsh, J. L., and McCarthy, B. J. (1973), *Biochem. Biophys. Res. Commun. 55,* 805.

Marsh, J. L., and McCarthy, B. J. (1974), *Biochemistry 13,* 3382.

Martin, M. A., and Hoyer, B. H. (1966), *Biochemistry 5,* 2706.

McCarthy, B. J., and Farquhar, M. N. (1972), in Evolution of Genetic Systems, Smith, H. H., Ed., New York, N.Y., Gordon and Breach, p 1.

Melchior, W. B., and Von Hippel, P. H. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 298.

Procunier, J. D., and Tartof, K. D. (1975), *Genetics 81,* 515.

Ptashne, M., Backman, K., Humayun, M. Z., Jeffrey, A., Maurer, R., Meyer, B., and Sauer, R. T. (1976), *Science 194,* 156.

Smith, M. J., Britten, R. J., and Davidson, E. H. (1975), *Proc. Natl. Acad. Sci U.S.A. 72,* 4805.

Sutton, W. D. (1972), *Nature (London) 237,* 70.

Tartof, K. D. (1971), *Science 171,* 294.

Thompson, W. F. (1976), *Plant Physiol. 57,* 617.

Ullman, J. S., and McCarthy, B. J. (1973), *Biochim. Biophys. Acta 294,* 416.

Walker, P. M. B., and McLaren, A. (1965), *J. Mol. Biol. 12,* 394.

Wetmur, J. G. (1976), *Annu. Rev. Biophys. Bioeng. 5,* 337.

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

Wilson, D. A., and Thomas, C. A., Jr. (1973), *Biochim. Biophys. Acta 331,* 333.