DNA Sequence Organization in the Pea Genome†

Michael G. Murray, Richard E. Cuellar,‡ and William F. Thompson*,‡

ABSTRACT: The reassociation kinetics of pea (*Pisum sativum* L.) DNA fragments (300 nucleotides) were measured with hydroxylapatite. The most slowly reassociating fragments do so with a rate constant of 2×10^{-4} L mol^{-1} s^{-1}, as determined from experiments with total DNA as well as with a tracer enriched for slowly renaturing sequences. This rate is about 1000 times slower than that observed for *Escherichia coli* DNA included as an internal kinetic standard, indicating a kinetic complexity of 4.5×10^9 nucleotide pairs or 4.6 pg of DNA per haploid nucleus. This estimate is in good agreement with previous chemical and cytophotometric measurements. The majority (85%) of the 300 nucleotide fragments contain repetitive sequences. While the reassociation of repetitive DNA could be modeled with two theoretical second-order components, the data did not specify a unique solution. The reassociation kinetics of isolated high- and low-frequency fractions indicate that repetitive sequence families in pea DNA probably cover a broad range of frequencies ranging from 100 to 10 000 or more copies per haploid genome. Single-copy sequences account for about 30% of the DNA, but because of extensive interpersion of repetitive sequences only about 15% of 300 nucleotide fragments reassociate with single-copy kinetics. From studies of hydroxylapatite binding as a function of fragment length, we conclude that the major class of single-copy sequences has a modal length of about 300 nucleotides. Long tracer reassociation kinetics indicate that sequences with an apparent repetition frequency of about 10 000 copies are interspersed at intervals of less than 1300 nucleotides throughout 75% of the genome. At a detection limit of about 3%, we find no single-copy sequences longer than 1000 nucleotides.

Studies on the linear arrangement of repetitive and single-copy sequences have contributed greatly to our understanding of complex eukaryotic genomes. A large number of animal species and a few higher plants have now been studied and most seem to conform to the same general pattern. In this pattern, short repetitive sequence elements (modal length about 300 ntp[1]) are interspersed throughout a large fraction of the single-copy DNA at intervals averaging from about 800 up to a few thousand nt, depending on the species (Davidson et al., 1975; Walbot & Goldberg, 1978). While there are a few species in which interspersion can be demonstrated only at much longer intervals (e.g., Crain et al., 1976), the basic pattern of short period interspersion is found in a remarkable variety of evolutionarily diverse organisms. Britten and Davidson (1969) have suggested that, in view of the nearly universal occurrence of this short period pattern, it may have some functional significance, perhaps in relation to the requirement for coordinate regulation of large numbers of unlinked genes in eukaryotic cells.

However, eukaryotes generally contain much more DNA than would appear to be required to account for a reasonable number of coding and regulatory functions (Hinegardner, 1976). Recent measurements also indicate that only a small fraction of the single-copy sequences may function as genes (Roshbash et al., 1975; Angerer et al., 1976; Galau et al., 1976). In a recent discussion of their regulatory model, Davidson et al. (1977) reviewed some of these observations and suggested that no more than a small fraction of the commonly observed short period interspersion pattern is likely to be composed of structural gene sequences with adjacent repetitive elements. In considering what significance might be attributable to the remaining DNA in these genomes, it will be useful to compare data from a wide variety of organisms, including those whose genome sizes and content of repetitive sequences differ greatly from one another, in order to obtain a better indication of the limits within which interspersion patterns may vary.

Plant genomes often contain more DNA, even on a haploid basis, than is found in most animal cells (Sparrow et al., 1972; Bennett & Smith, 1976; Hinegardner, 1976), and a larger fraction of plant DNA sequences are usually found to be repetitive (Flavell et al., 1974; Walbot & Goldberg, 1978). Relatively few studies of sequence interspersion have yet been carried out with plant DNA, however, and of those so far reported only the work of Flavell's group (Flavell & Smith, 1976; Smith & Falvell, 1977) has dealt with species having more than about 2 pg of DNA per haploid genome. Since the species studied by this group (wheat and rye) are closely related, it is not possible to come to any general conclusions concerning sequence organization in even moderately large plant genomes.

The garden pea (*Pisum sativum* L.) has a haploid (1C) DNA content of about 4.9 pg or 4.7×10^9 ntp (Bennett & Smith, 1976). Experiments described here show that single-copy sequences in pea DNA reassociate at a rate consistent with their representation only once per 4.5×10^9 ntp. Thus, the pea genome is physically as well as functionally diploid, although it contains approximately 2, 6, and 40 times as much DNA as the genomes of *Xenopus*, the sea urchin, and *Drosophila*, respectively (references in Davidson et al., 1975). Most of the pea genome consists of repetitive sequences which appear to be arranged into a broad spectrum of frequency classes between about 100 and 10 000 or more copies per ha-

† From the Carnegie Institution of Washington, Department of Plant Biology, Stanford, California 94305. *Received April 10, 1978.* Publication no. 630. This work was supported in part by National Science Foundation Grant PCM 7705656.

‡ Also, Department of Biological Sciences, Stanford University, Stanford, Calif. 94305.

[1] Abbreviations used are: nt, nucleotides; ntp, nucleotide pairs; C_0t, the product of molar concentration of DNA nucleotides and time of incubation (mol s L^{-1}); T_m, the temperature at which half of the nucleotides are unpaired; T_E, the temperature at which half of the DNA fragments elute from hydroxylapatite; Tris, tris(hydroxymethyl)aminomethane; Pipes, 1,4-piperazinediethanesulfonic acid; PB, an equimolar mixture of NaH_2PO_4 and Na_2HPO_4 (pH 6.8); EDTA, ethylenediaminetetraacetic acid.

0006-2960/78/0417-5781$01.00/0

ploid genome. Interspersion of these repeats among single-copy sequences in the pea genome is more extensive than previously reported for any other plant or animal.

Experimental Procedures

DNA Purification. Lyophilized light-grown leaf tissue is homogenized in 1.5 mL/g fresh weight of extraction buffer [50 mM Tris-HCl (pH 8.0 measured at 23 °C), 1% sodium dodecyl sulfate, 0.1% 2-mercaptoethanol, 10 mM EDTA, and 1 M $NaClO_4$] and 1.5 volumes of chloroform/octanol (24:1, v/v). After centrifugation, the aqueous phase is removed and the organic phase reextracted with 1 volume of extraction buffer. The combined aqueous phases are deproteinized once more with chloroform/octanol and then precipitated with 2.5 volumes of 95% ethanol. The ethanol precipitate is dissolved in 50 mM Tris-HCl (pH 8.0) and 10 mM EDTA. The resulting solution is then briefly dialyzed against the same buffer, clarified by low-speed centrifugation, and adjusted to 1% sarkosyl, 200 μg/mL ethidium bromide, and 49% (w/w) CsCl. After centrifugation (40 000 rpm for 40 h in a Beckman Type 65 rotor or 14 h in a Dupont-Sorval TV-850 rotor), the DNA band is visible under long wavelength UV illumination and can be easily removed with a large-bore syringe needle. Ethidium bromide is removed by partitioning the solution repeatedly against CsCl-saturated 2-propanol, and the CsCl then removed by dialysis against 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, and 0.1 M sodium acetate. Following digestion with DNase-free pancreatic RNase A (50 μg/mL for 1 h at 37 °C) and self-digested Pronase (100 μg/mL for 2 h at 37 °C), the DNA is further purified by hydroxylapatite chromatography in the presence of 8 M urea as described by Britten et al. (1969).

This procedure gives good yields of DNA, showing A_{260}/A_{230} ratios greater than 2.3 and A_{260}/A_{280} ratios close to 1.85. Purified preparations exhibited normal hyperchromicity upon thermal denaturation in phosphate buffer or 2.4 M tetraethylammonium chloride and did not accelerate the renaturation of bacterial standards (Murray & Thompson, 1977a).

DNA Shearing and Fragment Length Measurement. DNA was sheared to a modal single-strand fragment length of 300 nt in 0.2 M sodium acetate (pH 7.0), 5 mM EDTA, and 67% glycerol at −70 °C with a Virtis "60" homogenizer (60 000 rpm, 30 min). All DNAs were sized by alkaline agarose gel electrophoresis before use (McDonell et al., 1977; Murray & Thompson, 1977b). Molecular weight standards were obtained from phage λ DNA digested with restriction endonuclease *Eco*R1 (Allet et al., 1973) and from φX-174 RF DNA digested with *Hae*III (Sanger et al., 1977).

Preparation of Radioactively Labeled DNA Tracers. DNA was labeled in vitro using a modification of the system described by Dumas et al. (1971) for replicating single-stranded φX-174 DNA in vitro (Murray et al., 1977a). After removing foldback sequences introduced during in vitro replication, the tracer is indistinguishable by a number of criteria from in vivo labeled tracers that have also been stripped of foldback. In vivo labeled DNA was extracted from the roots of pea seedlings grown asepticly in the presence of [^{3}H]thymidine.

Preparation of Tracers of Different Sizes. Different sizes of radioactively labeled tracer molecules were prepared electrophoretically. ^{3}H-labeled copies were prepared in vitro from sheared templates of several different lengths and then subjected to preparative alkaline agarose electrophoresis. The gel was sliced and the DNA in each slice eluted electrophoretically (McDonell et al., 1977). Reported sizes for individual tracers were determined from analytical gel profiles *after* incubation under the reassociation conditions specified for each experiment.

Reassociation and Fractionation. Radioactive tracers were mixed with the desired driver DNA and dialyzed against 10 mM ammonium acetate. The mixtures were passed over small Chelex columns (Chelex 100, Bio-Rad), filtered through Metricel filters (Gelman GA-8), lyophilized, and dissolved in the desired reassocation buffer. Reassociation was carried out in 0.05, 0.12, or 0.4 M PB or in 1.0 M NaCl–0.01 M PB at 55, 60, 66, and 70 °C, respectively (T_m −25 °C). Samples containing 25 μg of unlabeled DNA and about 10 000 cpm of ^{3}H and/or 3000 cpm of ^{14}C-labeled tracer DNA were sealed in capillaries, denatured for 3 min at 108 °C, and incubated for varying lengths of time. Renaturation was terminated by quick cooling in ice and immediate dilution into excess cold 0.12 M PB.

Single-stranded DNA was separated from duplex-containing fragments by chromatography on hydroxylapatite (Bio-Rad HTP, lot no. 16020) as described by Britten et al. (1974). This lot of hydroxylapatite was tested before use to ensure that the fractionation protocol was sufficient to cleanly separate single-stranded and duplex-containing molecules (Martinson & Wagenaar, 1977). Because pea DNA forms extensive hyperpolymers, the duplex fraction was eluted with 0.12 M PB at 100 °C (Thompson, 1976). Reassociation of unlabeled DNA was assayed by measuring the A_{260} nm. Radioactivity in each fraction was measured by scintillation counting in a Triton–toluene cocktail (13.3 g of Omnifluor, Packard, 1 L of Triton X-100, and 2 L of toluene).

Labeled *E. coli* DNA was included as an internal standard in all reassociation mixtures. The kinetics of *E. coli* DNA renaturation were used to correct for the effects of viscosity on pea DNA reassociation (Wetmur & Davidson, 1968). Data are presented in the form of C_0t curves (Britten & Kohn, 1968). All C_0t values for buffers other than 0.12 M PB have been corrected to the equivalent C_0t for 0.12 M PB at 60 °C (Britten et al., 1974; Angerer et al., 1976). The results were analyzed in terms of theoretical second-order kinetic components using the computer program described by Pearson et al. (1977) and kindly provided by Dr. Eric Davidson.

Repeat Length Determination. Samples of >5000-nt-long in vivo labeled pea [^{3}H]DNA fragments in 1.0 M NaCl and 20 mM Pipes (pH 6.7) were reassociated at 70 °C to an equivalent C_0t value of 100. Single-stranded DNA was removed by digestion with nuclease S-1 (Miles Laboratories). Reassociated samples were adjusted to 25 mM sodium acetate (pH 4.4), 0.1 mM $ZnCl_2$, 150 mM NaCl, and 5 mM 2-mercaptoethanol (Britten et al., 1976) and digested for 45 min at 37 °C with either 5 or 50 units of enzyme/μg of DNA. A unit is defined as the amount of enzyme which renders 1 μg of single-stranded DNA acid soluble in 30 min at 37 °C. Five units of S-1/μg of DNA gave complete digestion of single strands to oligonucleotides (less than 50 nt) without noticeable degradation of native DNA when measured by gel electrophoresis. Digestion was terminated by the addition of an equal volume of 0.06 M PB containing 0.24 mM EDTA and 5 μg of unlabeled calf thymus DNA carrier, and the samples were immediately loaded onto hydroxylapatite columns at 25 °C. Columns were washed with 0.03 and 0.12 M PB at 25 °C, followed by 0.12 M PB at 60 °C prior to the elution of S-1 resistant duplexes with 0.4 M PB at 60 °C. DNA in this fraction was sized by neutral and alkaline agarose gel electrophoresis. Neutral gels contained 40 mM Tris–acetate (pH 8.1) and 5 mM EDTA.

Results

Reassociation of Total DNA. The reassociation kinetics of 300 nt fragments of total pea DNA are shown in Figure 1.

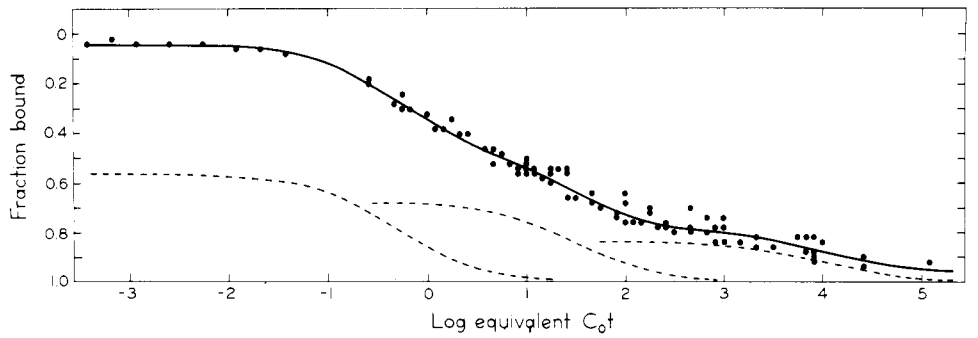

FIGURE 1: Reassociation kinetics of 300-nt pea DNA fragments. Samples were reassociated to various equivalent C_0t values and fractionated on hydroxylapatite as described under Experimental Procedures. C_0t values have been corrected to the equivalent C_0t values in 0.12 M PB at 60 °C (Britten et al., 1974). In vivo labeled pea [^{3}H]DNA fragments were used at both extremes of the curve, where small amounts of DNA in either single- or double-stranded fractions make accurate absorbance measurements difficult. *E. coli* tracers were included in all reactions to provide an internal rate standard to correct for rate retardation due to the viscosity of concentrated DNA solutions (see Figure 2). The solid line is a least-squares fit to the data using the three theoretical second-order components shown by dashed lines plus a 4% "very fast" fraction reacting prior to $C_0t = 0.01$. According to this model, 16% of the fragments reassociate with a single-copy rate constant of 0.0001 which becomes 0.0002 after correcting for viscosity effects. Forty-four percent of the fragments renature with a rate constant of 2.06, and 32% renature with a rate constant of 0.056 (after viscosity corrections). Four percent of the DNA was unreactable. The root mean square error of the indicated least-squares fit was 0.031.

Data are derived from four experiments in which reassociation was monitored by hydroxylapatite fractionation. Labeled *E. coli* DNA fragments were included in the pea DNA solutions to provide internal kinetic standards. At high concentrations (e.g., 5 mg/mL) of pea DNA, *E. coli* DNA reassociation was retarded by about twofold (see Figure 2 and Table I). This retardation is presumed to result from the high viscosity of concentrated DNA solutions (Wetmur & Davidson, 1968), an effect which might be enhanced in this case by the large hyperpolymers which form during reassociation of repetitive sequences in pea DNA (Thompson, 1976).

Four percent of the pea DNA fragments bind to hydroxylapatite at a C_0t value of about 5×10^{-4}, with no further binding being observed until C_0t values approach 0.1. The 4% initial binding will include any nonspecific adsorption to the hydroxylapatite, but probably results primarily from rapid *intramolecular* reassociation of inverted repeat or palindrome sequences (Huguet et al., 1975; Bazetoux et al., 1978). The remainder of the fragments react over some six decades of C_0t. Approximately 85% reassociate more rapidly than would be expected for sequences present only once per haploid genome and, thus, must contain a repetitive sequence element.

The data of Figure 1 were analyzed with a least-squares computer program designed to fit theoretical second-order components to the observed kinetics, and the elements of the simplest solution are illustrated in the figure. At least two theoretical components are required to describe the renaturation of repetitive sequences. Those illustrated in Figure 1 are (a) a "fast" (high frequency) component with a rate constant of 2.06 and accounting for 44% of the fragments and (b) a "slow" (low frequency) component with a rate constant of 0.033 and accounting for 32% of the fragments. After correction for viscosity effects in the more concentrated samples, the adjusted rate for the slow component is 0.056. The third component (16% of the fragments) reassociates about 1000 times more slowly than the *E. coli* standard.

Reassociation of Kinetically Fractionated DNA. We have prepared three fractions of the pea genome by procedures designed to separate the three putative frequency components of the fit illustrated in Figure 1. Pea DNA fragments (300 nt) were successively reassociated and fractionated on hydroxylapatite to yield the following fractions: "fast" repeats, reassociating between $C_0t = 0.01$ and 2; "slow" repeats, reassociating between $C_0t = 2$ and 200; and "single copy", reassociating after C_0t 1000. Radioactive tracers were prepared from each of these three fractions with the DNA polymerase I technique described above, and the reassociation kinetics of each tracer measured in the presence of a 2000-fold excess of unlabeled, 300-nt total DNA fragments.

TABLE I: Comparison of the Predicted and Observed Reassociation Kinetics for Fast and Slow Repeat Fractions.[a]

tracer	fraction		rate	
	predict.	obsd	predict.	obsd
fast	0.95	0.58	2.06	1.06
	0.05	0.42	0.06	0.05
slow	0.25	0.45	2.06	0.73
	0.75	0.55	0.06	0.03

[a] Fractions and rates are taken from the unconstrained least-squares fits shown in Figures 1 and 2. Rate constants have been normalized for viscosity effects as described in the text.

Figure 2 shows that all three tracer reactions are clearly heterogeneous, requiring a minimum of two theoretical components to achieve a satisfactory fit. In the case of the "single-copy" tracer, about 60% of the reactable label reassociates as a reasonably distinct component with a normalized rate constant of 2×10^{-4}. About 40% of the reactable tracer fragments reassociate with slow repeat kinetics. These fragments represent less than 6% of the total repetitive fraction.

Interpretation of the data for the two repetitive tracers is more complicated, since the results differ from those predicted by the simple model illustrated in Figure 1. Both tracers exhibit heterogeneous kinetics, requiring at least two major repetitive components for an adequate fit. The rate constants for the components fit to the tracer data differ somewhat from those predicted, even after viscosity corrections have been made. More importantly, however a comparison of predicted and observed two-component fits (Table I) indicates that there are many more slowly reassociating fragments in the "fast" tracer—and more rapidly reassociating fragments in the "slow" tracer—than would be expected from the model for two distinct repetitive components.

The tracer fragments (modal length 200 nt) were shorter than the 300-nt fragments used for preparative fractionation. Since the pea genome is characterized by extensive short period interspersion (see below), it is possible that some of the observed kinetic heterogeneity might be attributable to the separation in the tracers of fast and slow sequences which had been

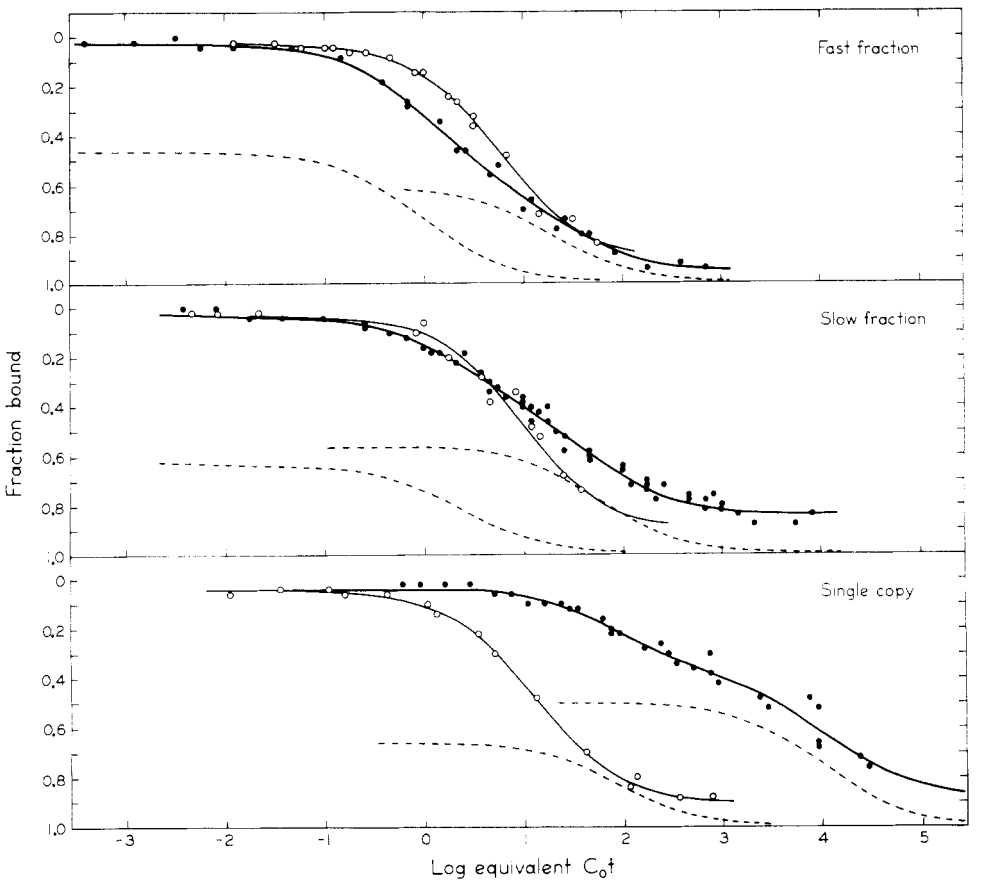

FIGURE 2: Reassociation of isolated kinetic fractions of the pea genome. 300-nt pea DNA fragments were separated into three fractions by sequential reassociation and fractionation on hydroxylapatite. The reassociation histories (Britten et al., 1974) and fractions of total DNA were: fast repeat, C_0t = 200 bound, C_0t = 2 bound, C_0t = 0.01 unbound (24%); slow repeat, C_0t = 200 bound, C_0t = 2 unbound (42%); single copy, C_0t = 200 unbound, C_0t = 1000 unbound (13%). ^{14}C-labeled tracers were prepared from each fraction as described under Experimental Procedures. Zero time binding DNA was removed by passage over hydroxylapatite immediately after denaturation. The resulting tracers had a mode length of 200 nt. Each tracer was mixed with a 2000-fold excess of unlabeled 300-nt fragments of total DNA. Reassociation and fractionation were carried out as described for Figure 1. In each panel, the heavy line represents the best two-component least-squares fit to the pea tracer data (●). The dashed lines represent the theoretical second-order components of each solution. The light lines (○) depict the reassociation of ^{3}H-labeled *E. coli* internal standards, with rate constants of 0.194, 0.114, and 0.091 in the "fast", "slow", and "single-copy" reactions, respectively. The "single-copy" tracer required a very slow repetitive component reassociating with a rate constant of 0.012 (0.026 after correcting for viscosity) in addition to the single-copy component; after correcting for viscosity, the single-copy rate is 0.0002. Root mean square errors for the fits to the pea tracer data were 0.023, 0.026, and 0.036 for the fast repeat, slow repeat, and single-copy fractions, respectively. Theoretical solutions for the fast and slow repeat fractions are given in Table I.

linked together on the same fragment when the preparative fractionation was carried out. However, while such an effect could lead to the appearance of more "slow" fragments in the "fast" tracer preparation, it would not explain the unexpectedly large fraction of "fast" fragments in the "slow" tracer. Furthermore, at C_0t = 2, the point at which the "fast" and "slow" fractions were originally separated, we find that 200-nt fragments of total DNA bind 85–90% as well as those 300-nt long. Thus, only 10–15% of the "fast" fraction would be expected to reassociate more slowly as 200-nt fragments than at the original length of 300 nt.

Thus, we must either conclude that repetitive sequence reassociation kinetics are more complex than the simple model would suggest or assume that our fractionation procedure somehow failed to resolve discrete components. To test the latter possibility, we have carried out model experiments in which mixtures of *E. coli* [^{14}C]DNA and [^{3}H]DNA from bacteriophage T4 were reassociated in the presence and absence of excess pea DNA fragments. The expected second-order kinetics and excellent resolution between these two different kinetic standards were obtained with or without excess pea DNA. Thus, complex structures formed during pea DNA reassociation (Thompson, 1976) did not affect the resolution. The preparative fractionation described above was scaled up directly from that used in these model experiments; thus, we believe our preparative procedure would have resolved discrete components of repetitive sequences similar to those of the model in Figure 1.

Thermal Stability of Repetitive DNA Fractions. In many eukaryotes it can be shown that some repetitive sequence families reassociate to form precisely paired duplexes, while others show significant amounts of sequence divergence and produce reassociated duplexes containing mismatched base pairs (Britten & Davidson, 1976; Bendich & Anderson, 1977; Walbot & Goldberg, 1978). Since base-pair mismatch reduces the rate of reassociation (Bonner et al., 1973; Hutton & Wetmur, 1973; Marsh & McCarthy, 1974), the presence of families with different degrees of sequence divergence could contribute to the heterogeneity of reassociation rate we observe among repetitive sequences in pea DNA. Few previous studies have directly addressed this possibility. However, Smith & Flavell (1975) have shown that at least part of the kinetic heterogeneity of wheat repetitive DNA could be explained by the effect of mismatch on reassociation.

To assess the extent to which mismatch effects might contribute to the kinetic heterogeneity of pea DNA, we measured

the thermal stability of duplexes formed at various C_0t values by total pea DNA and each of the repetitive tracer preparations. Reassociation was carried out at 35 °C below the T_m in order to maximize the formation of low-stability duplexes and, thus, enhance the resolution between precisely repeated and divergent families; hydroxylapatite thermal elution profiles were obtained in 0.075 M PB after removing single strands with 0.14 M PB at 50 °C. Preliminary experiments indicated that reducing the PB concentration to 0.075 M was necessary to avoid eluting some double strands before they were denatured (Martinson & Wagenaar, 1977). Labeled *E. coli* DNA was included in each sample during both reassociation and thermal elution. The elution profiles obtained for *E. coli* duplexes were virtually identical, with a single sharp peak and a T_E of 91 ± 0.7 °C (mean ± SD, six experiments).

Results of these experiments are shown in Figure 3. After reassociation to C_0t = 500, thermal elution profiles for the "fast" and "slow" tracers were virtually identical. Similar profiles were also obtained for the early and late reassociating portions of total DNA, using a subtraction procedure to determine the profile for sequences reassociated between C_0t values of 2 and 500. Early and late portions of the "fast" tracer also gave similar profiles. Significant differences in thermal stability were detected only between the early and late reassociating portions of the "slow" tracer. However, even in this case there is extensive overlap between two profiles, and the T_E of the late portion is only about 3 °C lower than that for the early portion. From the data of Bonner et al. (1973), we estimate that this small difference in average thermal stability would reduce the average reassociation rate in the late portion by 20% or less. Such a small discrepancy cannot explain the much larger heterogeneity seen in Figure 2.

Interspersion of Repetitive and Single-Copy DNA. Our studies of sequence interspersion use techniques developed by Davidson et al. (1973) and Graham et al. (1974) to exploit the fact that a renatured repetitive duplex region in an otherwise single-stranded molecule will cause the entire molecule to bind to hydroxylapatite. One can thus measure the fraction of DNA fragments of any given length which contain one or more regions of repetitive sequence, and from a series of such measurements it is possible to derive information concerning the spacing of repetitive elements interspersed among single-copy sequences.

Labeled DNA containing a distribution of fragment lengths was fractionated by preparative agarose gel electrophoresis in an alkaline system (McDonell et al., 1977; Murray & Thompson, 1977b) to yield tracers with sharper length distributions than those obtained by conventional shearing methods (Hinnebusch et al., 1978). When these tracers were analyzed on analytical gels, the width of the distributions at half-peak height averaged about 33% of the mean length, while tracers prepared by shearing techniques had bandwidths about 140% of the mean length. [Preliminary experiments with sheared tracers gave results generally similar to those presented here, although with much lower resolution (Murray et al., 1977b)]. Aliquots of each tracer were separately mixed with a 2000-fold excess of unlabeled 300-nt pea DNA fragments ("driver DNA"), reassociated to the desired C_0t value, and fractionated on hydroxylapatite. Similar aliquots of the same tracers (mixed with unlabeled calf DNA fragments to reduce nonspecific binding and handling losses) were fractionated immediately after denaturation (effective $C_0t < 10^{-4}$) to estimate the fraction binding at zero time as a result of intramolecular reactions. The fraction bound as a result of bimolecular reassociation was then calculated as described by Davidson et al. (1973).

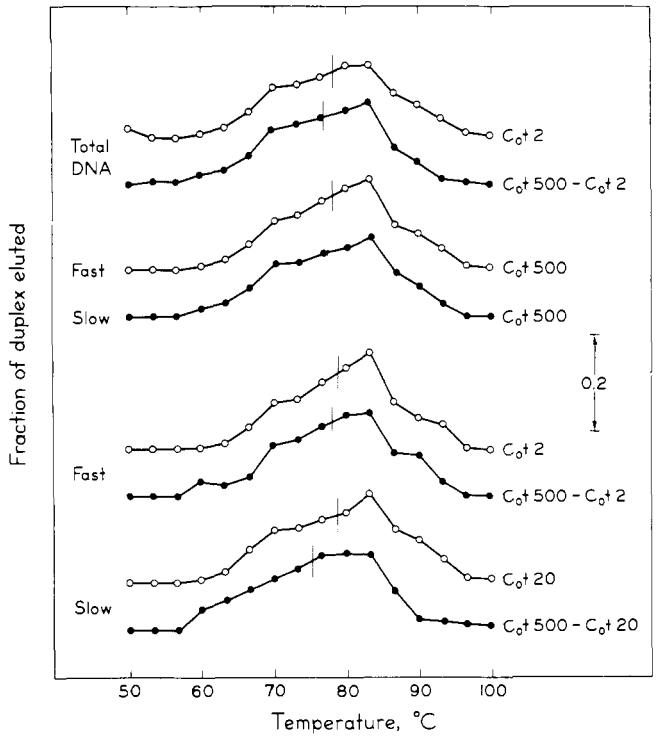

FIGURE 3: Thermal stability of reassociated "fast" and "slow" repeat tracers and total pea DNA. Aliquots of the tracers used in Figure 2 or total DNA tracer of the same size distribution were mixed with a 2000-fold excess of 300-nt-long unlabeled total pea DNA, and the mixtures were reassociated at T_m −35 °C. Thermal elution profiles were determined as described in the text at C_0t = 2 and 500 for total DNA and the "fast" fraction, and at C_0t = 20 and 500 for the "slow" fraction. Thermal profiles for the duplexes formed between the high and low C_0t values were determined by subtracting the fraction of the total applied (single and double stranded) DNA which eluted at a given temperature in the low C_0t profile from the fraction eluting at the same temperature in the high C_0t profile. Data for each temperature are presented as a fraction of the DNA reassociating between the indicated C_0t values; the scale is indicated by the bar at the right of the figure. The T_E for each curve is indicated by a vertical line.

Results from experiments conducted at two different C_0t values are shown in Figure 4. C_0t = 50 was chosen to provide a condition in which even long single-copy sequences would not reassociate but which would permit most (≥85%) of the repetitive DNA to react. Thus, the data shown in the lower curve of Figure 4 would normally be interpreted as reflecting the length distribution of single copy and slow repeat sequences which remain single stranded at this C_0t value (Davidson et al., 1973; Graham et al., 1974). However, this interpretation involves the assumption that the fraction of repetitive DNA reassociated at C_0t = 50 remains constant for tracers of widely different lengths, and such an assumption may be invalid where long regions or clusters of slowly reassociating DNA exist. In these cases, long tracer fragments containing such long regions would reassociate more rapidly than short fragments of the same sequence (Davidson et al., 1973; Hinnebusch et al., 1978), and the fraction reassociated at a given C_0t value would, therefore, increase with fragment length even in the absence of interspersion. This point can be clearly demonstrated with *E. coli* tracers in the inset to Figure 4. Such effects are likely to be insignificant where the amount of single-copy DNA is large in relation to the amount of repetitive DNA remaining unreacted at the chosen C_0t value, but they may become important when the fraction of single-copy sequences is small, as it is in many plants.

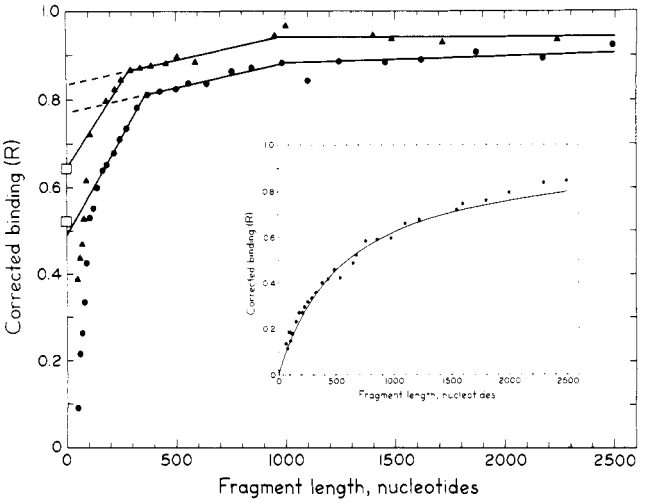

FIGURE 4: Hydroxylapatite binding of reassociated pea and *E. coli* DNA fragments of different lengths. In vitro labeled pea [^{3}H]DNA fragments of various lengths were prepared by alkaline agarose electrophoresis as described under Experimental Procedures. Labeled fragments were reassociated with a 2000-fold excess of unlabeled pea DNA (300 nt) to C_0t = 50 (●) or 680 (▲). The fraction bound to hydroxylapatite (corrected for zero time binding as described in the text) is plotted as a function of fragment length. To allow for thermal degradation, fragment lengths were determined after incubation to the appropriate C_0t value. The fraction of base paired nucleotides (□) was calculated from the hyperchromicity of 300-nt pea DNA fragments reassociated to C_0t = 50 or 680. Thermal denaturation was carried out in 0.12 M PB between 60 and 98 °C using a Gilford Model 2527 thermoprogrammer. The hyperchromicity of reassociated DNA (15.5 and 18.6% of the denatured A_{260} after incubation to C_0t = 50 or 680, respectively) was compared with that of native DNA (27.5%), allowing for the 2.5% hyperchromic effect observed for single strands under our conditions. The fraction base paired was 52% (= 15.5 −2.5/27.5 − 2.5) after incubation to C_0t = 50 and 64% (= 18.6 − 2.5/27.5 −2.5) after incubation to C_0t = 680. The binding of partially reassociated *E. coli* tracers as a function of fragment length is shown in the inset. In vivo ^{3}H-labeled *E. coli* fragments of various lengths were prepared electrophoretically, reassociated with excess driver DNA to C_0t = 3.4, and analyzed on hydroxylapatite as above. The solid line is a theoretical binding curve, calculated by assuming that tracer reassociation rates are proportional to fragment length (see text).

In the case of pea DNA at C_0t = 50, inspection of Figure 1 shows that the fraction of 300-nt fragments bearing unreacted repeats, although small in relation to the total repetitive DNA, is a large fraction of the total single-stranded material. To avoid ambiguity, therefore, we have conducted additional binding experiments in which reassociation was allowed to proceed to C_0t = 680, at which point virtually all of the repetitive sequences have reacted.

Taken together, the two sets of data provide no evidence for regions of single-copy DNA in excess of about 1000 nt. At C_0t = 680 the binding beyond 1000 nt remains essentially constant at 94%. This value is close to the expected maximum, since we have observed that about 4–6% of pea DNA fragments 1000-nt long bind to hydroxylapatite under these conditions only after a second reassociation with additional driver DNA. Similar observations have previously been interpreted as indicating that a small fraction of repetitive sequences fail to form hydroxylapatite-bindable duplexes during a single round of reassociation (Davidson et al., 1973). Thus, although the C_0t = 50 data alone might have been interpreted as indicating the presence of single-copy sequences in excess of 2500 nt long, a comparison of these data with that obtained at C_0t = 680 shows that the slow increase in binding above 1000 nt is best interpreted as reflecting length-dependent increases in reassociation of long repetitive regions. Independent evidence for the absence of a significant fraction of single-copy sequences longer than 1000 nt is given below.

Below 1000 nt, the binding curves can be divided into two approximately linear segments. We believe that the sharp break in slope which occurs at about 300 nt in the C_0t = 680 data indicates a major class of single-copy sequences in which repeated sequences are interspersed at intervals of about 300 nt, while increases in binding between 300 and 1000 nt reflect the presence of a smaller fraction of single-copy sequences with lengths up to 1000 nt. However, to reach this conclusion we must consider the possible contribution of single-copy DNA self-reassociation at C_0t = 680. For this purpose, we use the C_0t = 50 data to set upper limits on the amount of single-copy DNA in each length class and, thus, to derive upper limits for the extent of its self-reassociation.

By extrapolation to the Y axis from the linear segments of the C_0t curve, following the rationale of Graham et al. (1974), we estimate that up to about 11 and 25% of the pea genome *could* consist of single-copy DNA interspersed with repeats at intervals of about 1000 and 300 nt, respectively. These values are almost certainly overestimates, since they will include the effects of repetitive sequence reassociation similar to those seen in the C_0t = 50 data beyond 1000 nt. At C_0t = 680, fragments 300 and 1000 nt long and containing only single-copy DNA would be about 10 and 30% reacted, respectively. At 300 nt, therefore, we place the upper limit for binding due to single-copy self-reassociation at 3.6% [= 0.1 (25% + 11%)] and estimate that, at most, 2.2% [= (0.3 − 0.1)(11%)] additional single-copy reassociation would occur between 300 and 1000 nt. Since these maximum estimates are small in relation to the actual increases, we conclude that most of the observed C_0t = 680 binding reflects sequence interspersion.

Extrapolation to zero fragment length from the C_0t = 680 data between 150 and 300 nt suggests that 64% of the nucleotides are contained in repetitive sequences. Since our demonstration of a major class of 300-nt-long single-copy sequences depends on the validity of the binding data in this range, we compared the results obtained from the extrapolation procedure with independent estimates of the absolute repetitive sequence content derived from optical hyperchromicity measurements. After suitable corrections for the hyperchromicity of single strands, such measurements are known to provide an accurate measure of the fraction of nucleotides paired (Wetmur, 1976). After reassociation to C_0t = 680, we estimate that 64% of the nucleotides in the pea fragments are paired (Figure 4, legend). Assuming about 1% T_m reduction per 1% mismatch (Wetmur, 1976), we estimate that these repetitive duplexes contain between 5 and 7% mismatch. Thus, the optical estimate must be revised upward to about 70% to allow for mismatched bases within the repetitive duplexes. In addition, the extrapolated hydroxylapatite value must be corrected to 67% (= 64%/95%) because about 5% of the repetitive duplexes fail to bind to HAP (see above). We conclude that the binding data do accurately reflect sequence interspersion and that 67 to 70% of the nucleotides in the pea genome are contained in repetitive sequences.

Reassociation Kinetics of Long Tracers. Figure 5A shows an experiment in which labeled DNA fragments having a narrow length distribution around 1300 nt were reassociated with a 2000-fold excess of unlabeled 300-nt fragments of pea DNA. A large fraction (74%) of the long tracer fragments reassociate with kinetics suggesting a single, rapidly renaturing second-order component with a rate constant of about 1.0. The large increase in the fraction of fragments renaturing at this rate demonstrates that a major fraction of the pea genome contains elements of highly repeated families spaced at in-

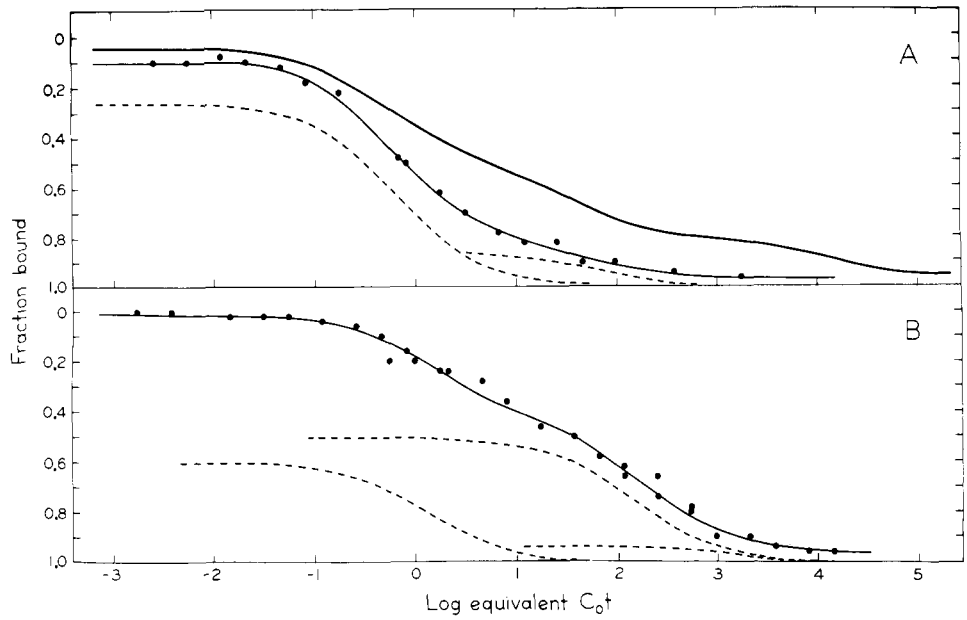

FIGURE 5: (A) Reassociation kinetics of 1300-nucleotide-long pea DNA fragments. ^{3}H-labeled long pea DNA fragments were prepared by alkaline agarose electrophoresis as described under Experimental Procedures and mixed with a 2000-fold excess of 300-nt unlabeled pea DNA and 1% *E. coli* [^{14}C]DNA to serve as an internal rate standard. Reassociation and fractionation were carried out as described in Figure 1. The upper curve shows the reassociation kinetics of 300-nt pea DNA fragments. The light line is an unconstrained fit to the 1300-nt tracer data (●) using two theoretical second-order components (dashed lines). Seventy-five percent of the long fragments reassociate with a rate of 1.05 after correcting for viscosity. Fourteen percent of the fragments reassociate with a rate of 0.012, and 4% of the DNA unreactable. The root mean square error of the indicated solution was 0.017. Similar analysis with 950-nt-long fragments gave similar results (data not shown). (B) Reassociation kinetics of 1000-nucleotide-long fragments remaining single stranded at C_0t = 10. 1000-nt-long in vivo ^{3}H-labeled pea fragments were reassociated to a C_0t value of 10, and those sequences that did not reassociate (18% of the total) were isolated by hydroxylapatite chromatography. These slowly reacting long fragments were mixed with a 700-fold excess of 300-nt unlabeled pea DNA and 1% *E. coli* [^{14}C]DNA to serve as an internal rate standard. Reassociation and fractionation were carried out as described in Figure 1. The upper curve (●) is a least-squares solution using three theoretical second-order components (dashed lines). For this fit, a single-copy rate was fixed at 0.0008 and the fraction of unreactable tracer was fixed at 4% (see text). In this model solution, 6% of the tracer, or 1% of the genome, reassociates with single-copy kinetics; the root mean square error was 0.024. An unconstrained fit to the same data could be obtained using only two components with no increase in error. In this solution, the slowest component accounts for 54% of the tracer (10% of the genome) and reassociates with a rate of 0.044.

vervals less than about 1300 nt. A small fraction of the long tracer reassociates more slowly, with a rate constant of 0.012. Single-copy sequences 1300-nt long would be expected to reassociate with a rate constant of 0.0009 (= 0.0002 × 1300/300) in a tracer/driver experiment such as this (Davidson et al., 1973). Since the slow component reassociates some 13 times faster than the maximum expected single-copy rate, we conclude that it reflects the reassociation of some of the less frequent repetitive elements rather than long single-copy sequences.

To probe further for the presence of long single-copy sequences, 1000-nt fragments that had not reassociated by a C_0t value of 10 were isolated, and their reassociation was followed in the presence of fresh unlabeled driver DNA. Since this tracer was enriched for slowly reassociating sequences by a factor of about 5, a small fraction of long single-copy DNA should be more readily detectable in this experiment. Figure 5B shows that a small component reassociating with the kinetics expected for 1000-nt-long single-copy DNA fragments can be fit to the data for this tracer fraction. If real, the component illustrated would account for 6% of the tracer or about 1% of the DNA in the pea genome. However, values for the rate constant of the slowest component and the fraction of unreactable fragments had to be fixed in order to obtain this solution. Unconstrained fits to the same data gave better solutions as judged by root mean square error and failed to produce evidence for any DNA fragments reassociating at the rate expected for single-copy sequences. A series of constrained and unconstrained fits indicates that between 0 and 15% of the tracer in Figure 5B might possibly reassociate with single-copy kinetics. Thus, we conclude that single-copy sequences longer than about 1000 nt constitute between 0 and 3% of the pea genome.

Length of Repetitive Sequence Elements. Having established that single-copy sequences in the pea genome are organized in a very short period interspersion pattern, we wished to determine whether or not repetitive sequences might be arranged in a similar fashion. Figure 6 illustrates the size distribution of S-1-resistant duplexes obtained after reassociating long fragments of pea DNA to an equivalent C_0t = 100. Digestions were carried out at both 5 and 50 units of S-1/μg of DNA in order to compare the results obtained after standard and more extensive digestion. In both cases, a large fraction of the resistant duplexes is quite short, with mode lengths of about 300–400 ntp. About 70 and 80% of the DNA are less than 1000 ntp long after the light and heavy digestions, respectively. The remainder of the duplexes form a broad distribution; there is no indication of a discrete class with a mode length above 400 ntp. Similar results were obtained when the lengths of the single strands in S-1-resistant duplexes were determined by alkaline agarose electrophoresis.

Discussion

Figures 1 and 2 show that single-copy sequences in total pea DNA reassociate about 1000 times more slowly than the *E. coli* DNA included as an internal standard, and thus the haploid pea genome would appear to contain about 1000 times as much DNA as does the *E. coli* chromosome. Since *E. coli* contains 4.5 × 10^6 ntp (Cairns, 1963; Klotz & Zimm, 1972), we estimate that the haploid pea genome contains about 4.5 × 10^9 ntp or 4.6 pg of DNA. A diploid (2C) nucleus would thus

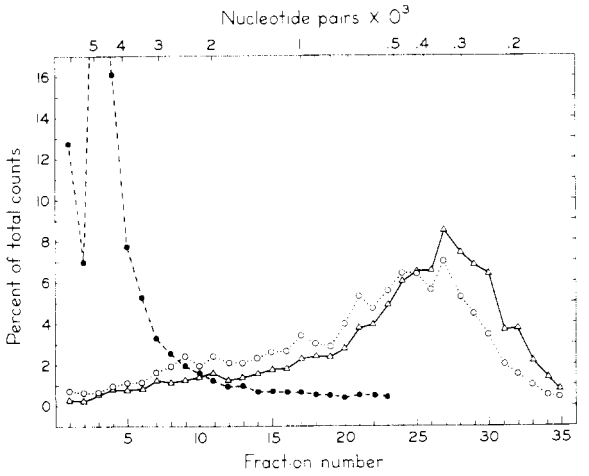

FIGURE 6: Size of the S-1-resistant repeated sequences in pea DNA. 5000-nt-long in vivo-labeled pea [^{3}H]DNA fragments in 1.0 M NaCl and 20 mM Pipes (pH 6.7) were denatured and reassociated at 70 °C to an equivalent C_0t = 100. After digestion with S-1 nuclease, the resistant duplexes were isolated by hydroxylapatite chromatography. The lengths of the resistant duplexes were determined by electrophoresis in neutral 1.5% agarose. The known fragments resulting from digestion of bacteriophase λ DNA with restriction endonuclease *Eco*R1 (Allet et al., 1973) and φX-174 DNA with restriction nuclease *Hae*III (Sanger et al., 1977) were used to derive the fragment length scale shown on the top. Profiles shown are for native DNA (●) and S-1-resistant repeated sequences after treatment with either 50 (○) or 5 (△) units of S-1/μg of DNA.

contain 9.2 pg according to our kinetic estimate. Van't Hoff & Sparrow (1963) measured an average DNA content of 11.7 pg/nucleus for cells in the terminal 2 mm of pea roots. If we assume from the results of Libbenga & Torrey (1973) that the cells in this region average 2.5C, the corrected value for 2C DNA content is 9.4 pg. From careful measurements of prophase nuclei, Bennett & Smith (1976) obtained estimates corresponding to 9.7 pg/2C nucleus. Our kinetic estimate agrees closely with both these values, indicating that the pea genome is physically as well as cytologically diploid.

Although our estimate for the kinetic complexity of pea DNA is similar to that mentioned in an early report by Sivolap & Bonner (1971), it differs dramatically from two other recent estimates (Pearson et al., 1978; Walbot & Goldberg, 1978). In both of the latter two cases, the slowest component identified reassociates at a rate several times faster than the single-copy rate determined from our data. However, Pearson et al. (1978) did not follow their reactions to completion, and no internal controls were included in either set of experiments. Factors capable of accelerating renaturation may be present in plant DNA preparations (Kemp & Merlo, 1975; Merlo & Kemp, 1976), and we have observed (Murray & Thompson, 1977a) that pea DNA purified by standard techniques may exhibit anomalous kinetics and accelerate the renaturation of internal standards. Therefore, all preparations of DNA used in our work were purified by banding in CsCl/ethidium bromide gradients, and the absence of acceleration was routinely verified by means of *E. coli* internal standards. These *E. coli* standards also permitted us to correct for viscosity effects in concentrated DNA solutions.

Approximately 15% of pea DNA fragments 300-nt long reassociate with the kinetics expected for single-copy sequences (Figure 1). However, hydroxylapatite measurements are known to underestimate the actual fraction of single-copy DNA when much of the genome is arranged in a short period interspersion pattern (Davidson et al., 1973). The data in Figure 4 suggest that about 30% of the pea genome is actually composed of single-copy sequences. From our estimate of 4.5 × 10^9 ntp per haploid genome, we may calculate that this amount of single copy DNA would contain about 1.4 × 10^9 ntp. Thus, the single copy fraction of the pea genome is more than ten times the size of the entire *Drosophila* genome. For comparison, the corresponding values for wheat and rye are 1.4 and 2 × 10^9 ntp/haploid genome (Flavell & Smith, 1976; Smith & Flavell, 1977), while those so far reported for other plant species range between 5 and 7 × 10^8 ntp on a haploid basis (Goldberg, 1978; Gurley et al., 1978; Kiper & Herzfeld, 1978; Walbot & Dure, 1976; Zimmerman & Goldberg, 1977).

The majority (about 70%) of the pea genome consists of repeated sequences of the moderate complexity "middle repetitive" type. Figure 1 shows that the repetitive sequences in pea DNA reassociate over about 5 decades of C_0t. There is clearly no discontinuity in the data of Figure 1 to constrain the possible solutions, and many different solutions can be obtained which fit the data nearly as well as that illustrated. Our attempts to fractionate repetitive sequences also failed to produce evidence for two discrete components. Some enrichment for more rapidly or slowly reassociating sequences was achieved, but both "fast" and "slow" fractions displayed markedly heterogeneous kinetics. Thermal stability profiles for duplexes formed at various points in the reaction of each fraction (Figure 3) do not differ from one another significantly, and thus very little of the observed kinetic heterogeneity can be attributed to participation of more highly diverged sequences or the formation of shorter duplex structures in the latter stages of either reaction. Thus, we believe that the heterogeneity we observe probably reflects the presence of repetitive families with a variety of different repetition frequencies between about 100 and more than 10 000 copies per haploid genome.

Relatively discrete frequency components which can be physically isolated and shown to reassociate with essentially ideal second-order kinetics do exist in several (but not all) animal genomes (e.g., Britten & Davidson, 1971; Hough & Davidson, 1972). Of the higher plant genomes so far analyzed in this way, only tobacco (Zimmermann & Goldberg, 1977) has been shown to contain a large fraction of repetitive DNA clearly belonging to a discrete frequency class. At present the available data are consistent with the hypothesis that the more highly repetitive higher plant genomes may be characterized by a more complex distribution of repetition frequency classes than that found in the genomes of typical animal species. This hypothesis would also be consistent with recent evidence that frequent amplification and reamplification events have played a major role in the evolution of the wheat genome (Flavell et al., 1977). However, even the most heterogeneous reassociation kinetics can usually be reconciled with as few as four or five theoretical components, and it is not possible to estimate the actual number of frequency classes with certainty for any genome. New approaches, such as the use of molecular clones representing individual repetitive sequences, will be required to resolve this question.

Data presented in Figures 4 and 5 show that the pea genome is characterized by extensive sequence interspersion at very short intervals. In most other eukaryotes so far examined, the shortest class of single-copy sequences averages from about one to a few thousand nt long, and a significant fraction appears to be much longer (Davidson et al., 1975; Walbot & Goldberg, 1978). In the case of peas, however, the predominant class of single copy sequences has a modal length of about 300 nt, and we cannot unambiguously detect any single-copy sequences longer than 1000 nt.

In genomes such as that of the pea where a clean separation

of repetitive and single-copy sequences at a single C_0t value is not possible, it is imperative to consider what effects different patterns of sequence organization might have on the results of binding experiments such as those in Figure 4. Where the fraction of slow repetitive sequences is large in relation to that of single copy DNA, as it is in the pea genome, results obtained at low C_0t values may reflect length-dependent changes in tracer reassociation kinetics as well as the effects of sequence interspersion. Thus, we base our conclusions primarily on the data obtained at C_0t = 680, where essentially all repetitive sequences should be reassociated, and use lower C_0t data only in deriving estimates for the maximum extent of single-copy reassociation. If we had used the C_0t = 50 data alone, we might have erroneously concluded that as much as 13% of the pea genome consists of single-copy sequences longer than 2500 nt.

Extensive short period interspersion is also revealed by the reassociation kinetics of long tracers (Figure 5). About 75% of all DNA fragments 1300-nt long reassociate with kinetics suggesting a single, highly repeated component. It is noteworthy that such a large fraction of this tracer should display apparently homogeneous kinetics, since the reassociation of repetitive DNA on 300-nt fragments is kinetically quite heterogeneous (see above). Extensive interspersion of high- and low-frequency repetitive sequences is required to explain these results. The large increase in the fraction of rapidly renaturing fragments in the long tracer indicates that elements of highly repeated sequence families are interspersed at intervals of 1300 nt or less in a major fraction of the pea genome.

An important conclusion from the experiments described in Figures 4 and 5 is that single-copy sequences longer than about 1000 nt do not constitute a readily measurable fraction of the pea genome. Most structural genes in animal genomes are single-copy sequences (Britten & Davidson, 1976; Goldberg et al., 1973), and the average structural gene is some 1200–1500 nt in length. Polysomal messenger RNA molecules from peas frequently have lengths greater than 1000 nt (Gray & Cashmore, 1976). Complementary DNA prepared from mRNA of tobacco cells (Goldberg et al., 1978) and pea buds (our unpublished observations) contains principally single-copy sequences. Thus, it is possible that mRNAs from peas might be constructed by splicing mechanisms similar to those recently demonstrated in animal systems (e.g., Breathnach et al., 1977; Brack & Tonegawa, 1977). However, since only a very small fraction of the DNA in a typical eukaryote may be required for structural genes (Galau et al., 1976; Britten & Davidson, 1976), it is also possible that coding sequences in a relatively large genome such as that of the pea plant may represent such a small fraction of the total DNA as to be undetectable in our experiments. To illustrate this point, we recall that a reasonable upper limit for pea single-copy DNA sequences longer than 1000 nt might be 3% of the genome. This 3% would contain more DNA than the entire *Drosophila* genome and enough coding capacity for about 10^5 average genes.

There is a remarkable similarity between the length distributions of repetitive and single-copy DNA sequence elements. In both cases, the dominant class is quite short—300 to 400 nt—and only a small fraction appears to be composed of longer sequence elements. Both classes of DNA in the pea genome, therefore, appear to have been subject to the same extensive very short period sequence interspersion. *Repetitive* sequences of this length are commonly observed in other eukaryotes (Davidson et al., 1975; Walbot & Goldberg, 1978), but a major class of single-copy sequences 300–400-nt long has been clearly observed before only in rye DNA (Smith & Flavell, 1977). In some cases, short single-copy sequences may not have been detected because too few data points were taken at short fragment lengths and/or because the length distribution of the tracers used was too broad to permit the required level of resolution. However, it is also unlikely that short single-copy sequences would be found to be quantitatively significant in genomes in which the repetitive fraction is much smaller than it is in pea DNA, since in such cases the principal result of randomly rearranging short segments of DNA would be a cryptic interspersion of single copy sequences with one another. On the other hand, where the majority of the genome consists of repetitive DNA, random rearrangements would juxtapose single-copy and repetitive elements much more frequently. Thus, a general process of very short period interspersion would produce the short single-copy sequences seen in pea DNA but only cryptic rearrangements of single-copy elements in less repetitive genomes.

Recent work from Flavell's laboratory (Smith et al., 1976; Flavell et al., 1977) has demonstrated the importance of sequence amplification and interspersion in the evolution of the wheat genome. Several major groups of repeated sequences have been identified which arose at various times during evolution. Sequences belonging to different groups have become extensively interspersed with one another in modern wheat. These observations suggest that amplification and translocation are fundamental processes in plant genome evolution. Frequent amplification, reamplification, and translocation could result in a complex distribution of repetition frequencies and would explain the extensive interspersion of short repetitive and single-copy sequences we observe in pea DNA.

Note Added in Proof

After this manuscript was prepared, it was reported (Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J., & Davidson, E. H. (1978) *Cell 14,* 899) that a sample of 26 cloned repetitive sequences from the sea urchin genome forms a continuous frequency distribution between about 3 and 12 500 copies per haploid genome. A similarly continuous distribution would be compatible with the kinetic heterogeneity we observe for repetitive sequences in pea DNA.

Acknowledgments

We thank Glenn Ford for his valuable assistance with computer programming.

References

Allet, B., Jeppesen, P. G. N., Katagiri, K. J., & Delius, H. (1973) *Nature (London) 241,* 120.

Angerer, R. C., Davidson, E. H., & Britten, R. J. (1976) *Chromosoma 56,* 213.

Bazetoux, S., Jouanin, L., & Huguet, T. (1978) *Nucleic Acids Res. 5,* 750.

Bendich, A. J., & Anderson, R. S. (1977) *Biochemistry 16,* 4655.

Bennett, M. D., & Smith, J. B. (1976) *Philos. Trans. R. Soc. London, Ser. B 274,* 227.

Bonner, T. I., Brenner, D. J., Neufeld, B. R., & Britten, R. J. (1973) *J. Mol. Biol. 81,* 123.

Brack, C., & Tonegawa, S. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 5652.

Breathnach, R., Mandel, J. L., & Chambon, P. (1977) *Nature (London) 270,* 314.

Britten, R. J., & Davidson, E. H. (1969) *Science 165,* 349.

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

Britten, R. J., & Davidson, E. H. (1971) *Q. Rev. Biol. 46,*

111.
Britten, R. J., & Kohne, D. E. (1968). *Science 161,* 529.
Britten, R. J., Pavich, M., & Smith, J. (1969) *Carnegie Inst. Washington, Yearb. 68,* 400.
Britten, R. J., Graham, D. E., & Neufeld, B. R. (1974) *Methods Enzymol. 29,* 363.
Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M., & Davidson, E. H. (1976) *J. Mol. Evol. 9,* 1.
Cairns, J. (1963) *Cold Spring Harbor Symp. Quant. Biol. 28,* 43.
Crain, W. R., Davidson, E. H., & Britten, R. J. (1976) *Chromosoma 59,* 1.
Davidson, E. H., Hough, B. R., Amenson, C. S., & Britten, R. J. (1973) *J. Mol. Biol. 77,* 1.
Davidson, E. H., Galau, G. A., Angerer, R. C., & Britten, R. J. (1975) *Chromosoma 51,* 253.
Davidson, E. H., Klein, W. H., & Britten, R. J. (1977) *Dev. Biol. 55,* 69.
Dumas, L. B., Darby, G., & Sinsheimer, R. L. (1971) *Biochim. Biophys. Acta 228,* 407.
Flavell, R. B., & Smith, D. B. (1976) *Heredity 37,* 231
Flavell, R. B., Bennett, M. D., Smith, J. B., & Smith, D. B. (1974) *Biochem. Genet. 12,* 257.
Flavell, R. B., Rimpau, J., & Smith, D. B. (1977) *Chromosoma 63,* 205.
Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., & Davidson, E. H. (1976) *Cell 7,* 487.
Goldberg, R. B. (1978) *Biochem. Genet. 16,* 45.
Goldberg, R. B., Galau, G. A., Britten, R. J., & Davidson, E. H. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70,* 3516.
Goldberg, R. B., Hoscheck, G., Kamalay, J. C., & Timberlake, W. E. (1978) *Cell 14,* 123.
Graham, D. E., Neufeld, B. R., Davidson, E. H., & Britten, R. J. (1974) *Cell 1,* 127.
Gray, R. E., & Cashmore, A. R. (1976) *J. Mol. Biol. 108,* 595.
Gurley, W. B., Hepburn, A. G., & Key, J. L. (1978) *Biochim. Biophys. Acta* (in press).
Hinegardner, R. (1976) in *Molecular Evolution* (Ayala, J., Ed.) p 179, Sinauer Associates, Inc., Sunderland, Mass.
Hinnebusch, A. G., Clark, V. E., & Klotz, L. C. (1978) *Biochemistry 17,* 1521.
Hough, B. R., & Davidson, E. H. (1972) *J. Mol. Biol. 70,* 491.
Huguet, T., Jouanin, L., & Bazetoux, S. (1975) *Plant Sci. Lett. 5,* 379.
Hutton, J. R., & Wetmur, J. G. (1973) *Biochemistry 12,* 558.
Kemp, J. D., & Merlo, D. J. (1975) *Biochem. Biophys. Res. Commun. 67,* 1522.
Klotz, L. C., & Zimm, B. H. (1972) *J. Mol. Biol. 72,* 779.
Kiper, M., & Herzfeld, F. (1978) *Chromosoma 65,* 335.
Libbenga, K. R., & Torrey, J. G. (1973) *Am. J. Bot. 60,* 293.
Marsh, J. L., & McCarthy, B. J. (1974) *Biochemistry 13,* 3382.
Martinson, H. G., & Wagenaar, E. B. (1977) *Biochem. Biophys. Acta 474,* 445.
McDonell, M. W., Simon, M. N., & Studier, F. W. (1977) *J. Mol. Biol. 110,* 119.
Merlo, D. J., & Kemp, J. D. (1976) *Plant Physiol. 58,* 100.
Murray, M. G., & Thompson, W. F. (1977a) *Carnegie Inst. Washington, Yearb. 76,* 255.
Murray, M. G., & Thompson, W. F. (1977b) *Carregie Inst. Washington, Yearb. 76,* 259.
Murray, M. G., Belford, H. S., & Thompson, W. F. (1977a) *Carnegie Inst. Washington Yearb. 76,* 262.
Murray, M. G., Preisler, R. S., & Thompson, W. F. (1977b) *Carnegie Inst. Washington Yearb. 76,* 240.
Pearson, W. R., Davidson, E. H., & Britten, R. J. (1977) *Nucleic Acids Res. 4,* 1727.
Pearson, W. R., Smith, S. L., Wu, J. J., & Bonner, J. (1978) *Plant Physiol. 62,* 112.
Roshbash, M., Campo, M. S., & Gummerson, K. S. (1975) *Nature (London) 258,* 682.
Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchinson, C. A., Slocombe, P. M., & Smith, M. (1977) *Nature (London) 265,* 687.
Sivolap, Y. M., & Bonner, J. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68,* 387.
Smith, D. B., & Flavell, R. B. (1975) *Chromosoma 50,* 223.
Smith, D. B., & Flavell, R. B. (1977) *Biochim. Biophys. Acta 474,* 82.
Smith, D. B., Rimpau, J., & Flavell, R. B. (1976) *Nucleic Acids Res. 3,* 2811.
Sparrow, A. H., Price, H. J., & Underbrink, A. G. (1972) *Brookhaven Symp. Quant. Biol. 23,* 451.
Thompson, W. F. (1976) *Plant Physiol. 57,* 617.
Van't Hoff, J., & Sparrow, A. H. (1963) *Proc. Natl. Acad. Sci. U.S.A. 49,* 897.
Walbot, V., & Dure, L. S. (1976) *J. Mol. Biol. 101,* 503.
Walbot, V., & Goldberg, R. (1978) in *Nucleic Acids in Plants* (Davies, J. W., & Hall, T., Eds.) in press, Chemical Rubber Company Press, Cleveland, Ohio.
Wetmur, J. G. (1976) *Annu. Rev. Biochem. Bioeng. 5,* 337.
Wetmur, J. G., & Davidson, N. (1968) *J. Mol. Biol. 31,* 349.
Zimmerman, J. L., & Goldberg, R. B. (1977) *Chromosoma 59,* 227.