Cell, Vol. 3, 315–325, December 1974, Copyright © 1974 by MIT

A System for Mapping DNA Sequences in the Chromosomes of Drosophila melanogaster

Pieter C. Wensink,* David J. Finnegan, John E. Donelson,* and David S. Hogness†
Department of Biochemistry
Stanford University School of Medicine
Stanford, California 94305

Summary

Individual segments of the chromosomal DNA in D. melanogaster were isolated, and the sequences they contain were analyzed for repetition and mapped within the polytene chromosomes. Isolation was achieved by first constructing circular hybrid DNA molecules consisting of single chromosomal segments linked by poly(dA):poly(dT) joints to single molecules of the tetracycline resistance plasmid, pSC101. Tetracycline-sensitive E. coli were transformed to resistance by this heterogeneous population of hybrid molecules and homogeneous populations of different hybrids were isolated from the clones of transformants. Three hybrid plasmids (pDm 1, 2, and 4) were studied in detail. Each exhibits the structure expected from the method of construction and none exhibits internal sequence repetition detectable by reassociation kinetics. The D. melanogaster sequences in pDm2 and 4 belong to a single class defined by little or no repetition within the genome and localization to a single chromomeric region in the polytene chromosomes. The characteristics of this class, which also includes 4 of a second set of 6 hybrids, are not compatible with tandem repetition models for the chromomere. The sequences in pDm1 are repeated 90 times and are located in 15 different chromomeric regions and within the chromocentric β-heterochromatin. This distribution is of the kind predicted by certain regulatory models, for example, that of Britten and Davidson (1969).

Introduction

The DNA in the euchromatic arms of Drosophila melanogaster chromosomes is subdivided into approximately 5000 chromomeric units. These units are recognized by the bands they form in polytene chromosomes, and are individually defined as the segment of a chromosomal DNA molecule that is included in a band plus an adjacent interband. A major problem concerning the arrangement of genetic information in eucaryotic chromosomes is posed by two contrasting properties of these units. Cytogenetic experiments indicate that the number of complementation groups, or cistrons, per chromomeric unit is very small, perhaps only one (Judd, Shen, and Kaufman, 1972; Hochman, 1973). By contrast, the DNA content of the individual units is large (Beerman, 1972; Rudkin, 1972), exhibiting a mean of 26 kb and a range that extends to more than 100 kb [kb (kilobases) is a unit of length equal to 1000 bases or base pairs in single- or double-stranded nucleic acids, respectively]. The average unit in D. melanogaster thus contains enough DNA to code for 26 different polypeptides of average size (Laird, 1973), and some units could code for more than 100. Why is the potential genetic information allowed by the DNA content so much greater than that indicated by the number of detected cistrons? Is the structure of the eucaryotic cistron quite different from that in procaryotes, where one cistron generally codes for a single polypeptide and the DNA content is consistent with the coding ratio?

A large number of models have appeared that offer possible answers to these questions [see Beermann (1972) and Bishop (1974) for reviews]. Most of these models can be distinguished by the topography that each imposes on repetitive and single-copy sequences, and on transcribed and translated sequences in the chromomeric DNA. The subject of this paper is the development of an experimental system whereby the different models can be tested by mapping the relevant sequences within and among the chromomeric units.

A prerequisite for this sequence mapping is the isolation of homogeneous populations of individual segments of D. melanogaster DNA. This has been accomplished by first constructing hybrid DNA molecules, each of which consists of a segment of D. melanogaster DNA linked to the DNA of an E. coli plasmid (Figure 1). This link was formed by adding short poly(dA) or poly(dT) tails to the 3′-OH termini of the respective DNAs, and then annealing these tails to form circular hybrid molecules (Jackson, Symons, and Berg, 1972; Lobban and Kaiser, 1973). The tetracycline resistance plasmid, pSC101, was chosen because DNA inserted at its single EcoR1 cleavage site does not prevent its autonomous replication or the expression of its tetracycline resistance gene(s), (Cohen et al., 1973; Morrow et al., 1974). Transformation of tetracycline-sensitive E. coli to resistance by this heterogeneous population of hybrid DNAs yielded clones of transformants from which the desired homogeneous populations of hybrid DNA molecules were isolated.

In this paper we examine three hybrid DNAs that were isolated and amplified by this method. Each consists of a single segment of D. melanogaster DNA linked to a single pSC101 DNA by

*Present address: P. C. W.: Rosenstiel Center, Brandeis University, Waltham, Massachusetts 02154; J. E. D.: Department of Biochemistry, School of Medicine, University of Iowa, Iowa City, Iowa 52240.
†To whom reprint requests should be sent.

poly(dA):poly(dT) joints. The sequences in the D. melanogaster segment have been analyzed with respect to the number of times they are repeated within the segment and within the D. melanogaster genome. They have also been mapped by in situ hybridization to polytene chromosomes (Pardue, Brown, and Birnsteil, 1973).

Results

Construction and Isolation of Hybrid DNAs

The procedure used to construct the hybrid DNAs is a modification of previously described methods (Jackson et al., 1972; Lobban and Kaiser, 1973) and is given in Figure 1. When products of the annealing reaction from each of several experiments were analyzed in the electron microscope, the fraction of molecules exhibiting the expected circular structure fell in the range, 16 ($\pm$ 9) %. This fraction dropped to $\leq$ 1% in control experiments where either the pSC101 or the D. melanogaster DNA was left out of the annealing mixture. The mean length of the circles produced from the mixture of the two DNAs equaled the sum of their mean lengths. The remaining molecules in the annealed population were mostly linear, but also included the branched forms observed by Lobban and Kaiser (1973), (see legend to Figure 1).

Transformation of tetracycline-sensitive h303 cells to resistance by the annealed mixture occurred with an efficiency approximately two orders of magnitude less than that observed with closed circular pSC101 DNA (10^{-6} transformants/molecule) under our conditions (legend to Figure 1). When the contour lengths of the circular plasmid DNAs isolated from each of nine transformants were determined by electron microscopy, three (pDm1, pDm2, and pDm4) exhibited lengths significantly greater than pSC101 DNA (Table 1) and were selected for further analysis. These three presumptive hybrid DNAs are as homogeneous with respect to length as is the pSC101 DNA [Footnote (a) in Table 1].

The Expected and Observed Structures of pDm DNAs are Equivalent

A series of experiments was performed to determine whether the pDm plasmids exhibit the structure expected from the method of construction. Heteroduplexes were formed between pSC101 and each pDm plasmid by denaturing and annealing a mixture of the two randomly nicked DNAs (Figure 2). In each case, heteroduplexes were observed that consist of a double-stranded ring the size of pSC101 DNA and an attached single-stranded ring, with a length equal to the difference between the lengths of the pDm and pSC101 DNAs (Table 1).

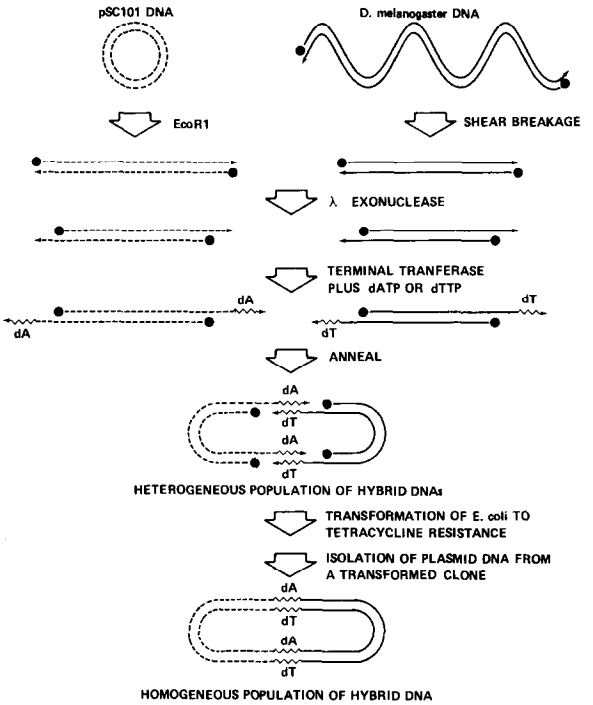

Figure 1. Construction and Isolation of pSC101–D. melanogaster Hybrid DNA

Arrowheads and dots represent the 3′ and 5′ termini of the strands, respectively. Closed circular pSC101 DNA was converted to linear molecules (99.5% conversion; electron microscopic assay) by cleavage with EcoR1 endonuclease (Mertz and Davis, 1972). D. melanogaster DNA, freed of most satellites by sedimentation in actinomycin D/CsCl gradients (Peacock et al., 1973), was sheared to a mean length of 8 kb by stirring (Hogness and Simmons, 1964) for 30 min at 1500 rpm and 0°C. Hybrid DNA molecules were formed from these two populations according to a modification of the method of Jackson et al. (1972) and Lobban and Kaiser (1973). An average of 25 nucleotides per 5′ terminus was removed from each DNA with λ exonuclease, and an average of about 40 dT or dA nucleotides per 3′ terminus were then added to the respective DNAs with calf thymus terminal transferase under the conditions of Lobban and Kaiser (1973), except that the temperature was lowered to 25°C. The poly(dT)-terminated D. melanogaster DNA (1.2 μg/ml) was annealed to the poly(dA)-terminated pSC101 DNA (0.6 μg/ml) in 0.1 M NaCl in TE for 10 min at 65°C and then for 2 hr at 45°C. Electron microscopic examination of 900 molecules revealed that the DNA preparation used to isolate the three hybrid DNAs described in the text consisted of 8.6% circles and 77% linears; the remainder consisted of unscorable tangles (7.6%) and branched forms (7.1%; these include branched linear and circular molecules).

Transformation was effected by the following modification of the method of Mandel and Higa (1970). Strain h303 in exponential growth at 37°C in L broth (Lennox, 1955) lacking glucose was chilled when the absorbance at 600 nm = 0.6, resuspended in an equal volume of 10 mM $MgSO_4$, held at 0°C for 30 min, and resuspended twice in 50 mM $CaCl_2$—first in 1/2 the original volume, and 15 min later (0°C) in 1/10 that volume. These cells (0.2 ml) were mixed with 0.1 ml of DNA in 10 mM each of $MgCl_2$, $CaCl_2$ and Tris-HCl(pH 8.1), held 15 min at 0°C, 2 min at 42°C, and 10 min at 23° $\pm$ 3°C prior to mixing with 15 ml of tetracycline supplemented (15 μg/ml) soft L agar (0.8%) on the Petri plate. Easily visible tetracycline-resistant colonies appear in the agar after 36–48 hr at 37°C.

Table 1. Lengths and Buoyant Densities of the Plasmids and Their Segments

Plasmid	Plasmid DNA[a] (kb)	Heteroduplex DNA[b] Double-strand (kb)	Heteroduplex DNA[b] Single-strand (kb)	Plasmid density[c] (g/cc³)	ρ_{Dm}[d] (g/cc³)
pSC101	9.22 ± 0.06			1.7104	
pDm1	12.09 ± 0.11	9.3 ± 0.2	3.1 ± 0.1	1.7071	1.697
pDm2	17.95 ± 0.16	9.2 ± 0.2	8.9 ± 0.3	1.7046	1.699
pDm4	21.84 ± 0.19	9.3 ± 0.2	13.0 ± 0.2	1.7082	1.707

[a] Lengths of 51–95 sample and reference molecules were measured from aqueous spreads. Standard deviations (s) for pSC101, pDm1, 2 and 4 are, respectively, 0.34, 0.34, 0.58, and 0.62 kb; $(s)/(m)^{0.5}$ for each DNA (where m = mean length) falls in the range 0.118 (± 0.020) $kb^{0.5}$, indicating that the DNAs are equivalently homogeneous with respect to length (Davis et al., 1971).

[b] Heteroduplexes were formed and spread for electron microscopy as described in Figure 2; 11–20 heteroduplexes and reference molecules were measured in each case.

[c] Buoyant densities were determined in CsCl as described by Champoux and Hogness (1973) and are relative to E. coli DNA at 1.7100 g/cc³.

[d] ρ_{Dm} is the buoyant density in CsCl of the D. melanogaster segment in each pDm DNA, and was calculated from

$$\rho_{Dm} = (\rho_{pDm})/[1 + (L_{pSC}/L_{Dm})(\Delta\rho/\rho_{pSC})]$$

where ρ_{pDm} and ρ_{pSC} are the buoyant densities of the pDm and pSC101 plasmids, $\Delta\rho = \rho_{pSC} - \rho_{pDm}$, and L_{pSC} and L_{Dm} are the lengths of the pSC101 DNA and the D. melanogaster segment in the pDm DNA. L_{Dm} is taken as the difference between the pDm and pSC101 plasmid lengths. See Champoux and Hogness (1972) for the method of deriving this equation.

λ DNA (46.5 kb; Davidson and Szybalski, 1971) was used as the reference for pSC101, and pSC101 then used as reference for the pDm DNAs and for the double-stranded segment of the heteroduplexes; M13 DNA (6.6 kb; Marvin and Hohn, 1969) was used as the single-stranded reference. The standard errors given above include all measurements required to relate the sample length to the primary references, λ and M13.

The circular pDm DNAs are therefore divided into two segments: one containing the base sequence of a single pSC101 genome, and the other containing a foreign base sequence, which, as will become apparent, derives from the D. melanogaster genome. The presence of a D. melanogaster segment in the pDm DNAs is indicated by their buoyant densities in CsCl. Table 1 shows that these densities are consistently less than that of pSC101, and that the calculated densities of the D. melanogaster segment, ρ_{Dm}, are well within the range exhibited by the main band of D. melanogaster DNA (Schachat and Hogness, 1973).

If the segment of D. melanogaster DNA was inserted at the EcoR1 cleavage site of pSC101 (Figure 1), then substitution of EcoR1-cleaved pSC101 DNA for the randomly nicked form in the above reaction should change the topology of the heteroduplex from two linked rings to a single ring such as that diagrammed in Figure 4. If insertion is elsewhere, both kinds of heteroduplexes should exhibit the two-ring structure. When this substitution was made, all pDm plasmids formed the expected one-ring heteroduplex, each consisting of a single-stranded and a double-stranded segment with lengths approximately those given in Table 1 for the corresponding regions in the two-ring heteroduplexes.

Insertion of D. melanogaster DNA at the EcoR1 cleavage site of pSC101 by the formation of two poly(dA):poly(dT) joints should alter the base sequence at this site so that it cannot be cleaved by the enzyme. A given pDm DNA therefore should be cut by EcoR1 only if sites exist in the D. melanogaster segment, in which case at least one fragment should be longer than pSC101 DNA, and it is unlikely that any fragment will exhibit the length of this DNA. These expectations were confirmed by electron microscopic observation of each pDm DNA after treatment with EcoR1. pDm2 yielded equimolar amounts of four fragments, with mean lengths and standard errors of 14.43 ± 0.09, 1.92 ± 0.02, 0.99 ± 0.02, and 0.44 ± 0.02 kb (N = 48–51); pDm4 yielded two equimolar fragments of 13.66 ± 0.07 and 8.23 ± 0.07 kb (N = 41 and 40); and no fragments were observed from pDm1 (<2% cleavage) under conditions where greater than 98% of the SV40 internal reference was cleaved (Morrow and Berg, 1972).

We next ask whether the poly(dA):poly(dT) joints are retained through the transformation and amplification steps. The largest EcoR1 fragment from pDm2 and 4 should contain the pSC101 segment and both joints. We tested this expectation for pDm2 by examining the heteroduplexes formed between EcoR1-cleaved pDm2 and pSC101 DNAs. When spread in 40% formamide, these heteroduplexes exhibit the linear structure diagrammed in the left half of Figure 3. If pDm2 retains both joints, each of the complementary heteroduplexes will contain the indicated poly(dA) and poly(dT) segments, and these may pair to circularize the heteroduplex. Circular heteroduplexes were not seen when spread in 40% formamide at room temperature (23° ± 3°C), but were frequently observed in aqueous spreads, where the single-stranded tails collapse to form a "blob" (Figure 3). This instability under the mildly denaturing conditions of formamide spreading suggests that the structure responsible for closure is either a small poly(dA):poly(dT) duplex or consists of randomly paired bases in the two tails.

The first, but not the second of these mechanisms for closure should be inhibited by the addition of dA residues to the 3′-OH terminus of the pSC101 strand with E. coli DNA polymerase I, which can use the poly(dT) in the pDm2 strand as a template

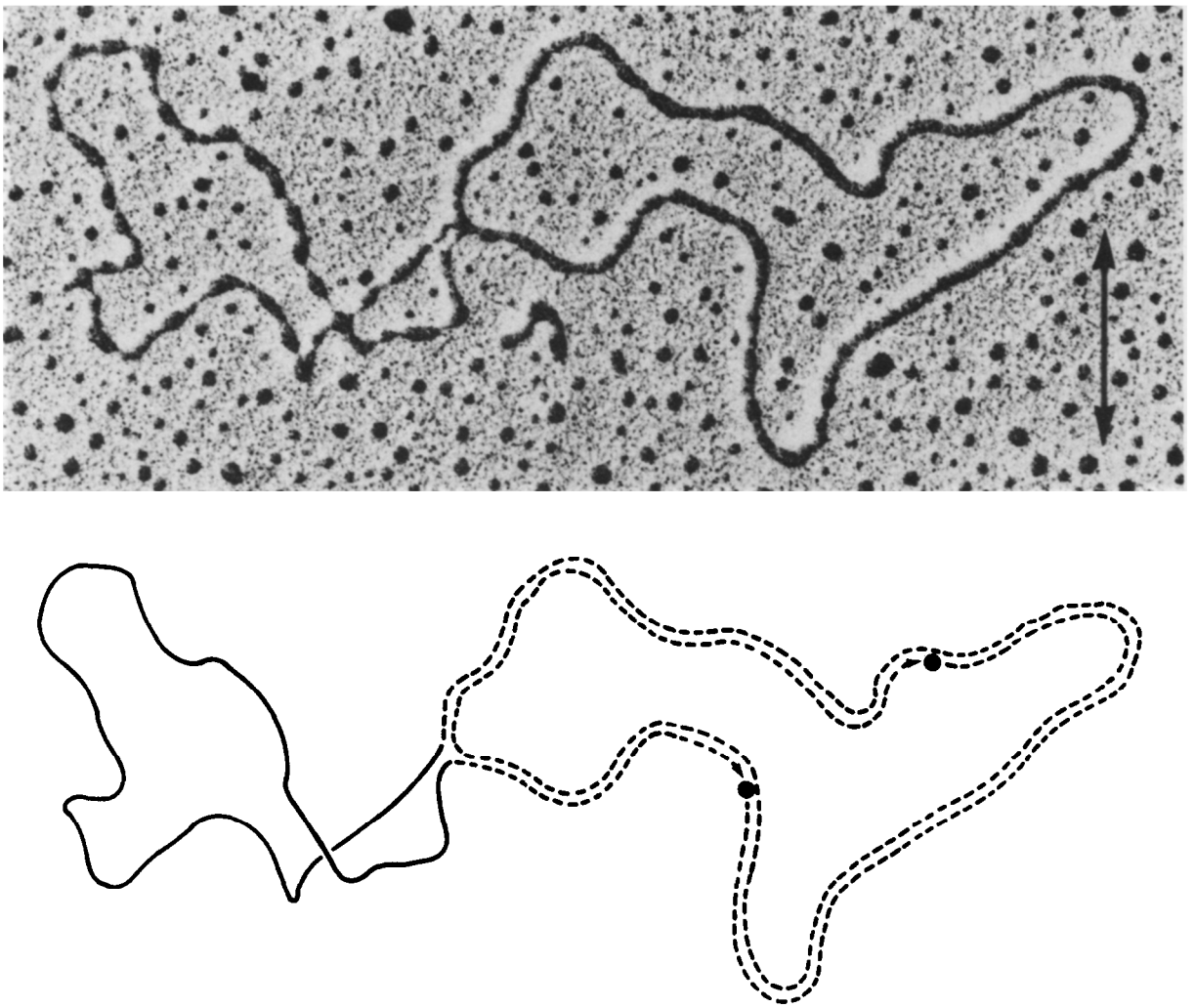

Figure 2. Heteroduplex Between pDm2 and pSC101

The circular double-stranded pDm2 and pSC101 DNAs each contained an average of one single-stranded break per molecule prior to mixing. These breaks occurred spontaneously after isolation of the closed circular DNA and are presumed to be randomly distributed. Denaturation and formation of heteroduplexes in 47% formamide were performed as described by Davis et al. (1971). The symbols in the diagram (arrowheads, etc.) have the same meaning as in Figure 1. Heteroduplexes were spread in 40% formamide for electron microscopy, and the double-arrowed line in the micrograph represents 1.0 kb (double-stranded) or 1.2 kb (single-stranded).

(Figure 3). Inhibition of closure was observed when the heteroduplexes were treated with DNA polymerase and dATP, but not when the dATP was absent or replaced by dTTP, dGTP, or dCTP (Table 2). The simplest explanation for these observations is that both of the poly(dA):poly(dT) joints are retained in pDm2.

The experiment illustrated in Figure 4 indicates that at least one of these joints is retained in pDm1. Poly(dA) extensions of mean length = 0.21 kb were added to the 3′ termini of the linear phage P22 DNA by the method described in Figure 1, and these molecules were used as a probe for a single-stranded poly(dT) segment in the one-ring heteroduplexes formed between randomly nicked pDm1 and EcoR1-cleaved pSC101 DNAs. The expected result of pairing between the poly(dA) probe and such a poly(dT) segment is a lariat, which in aqueous spreads should consist of a tail the length of P22 DNA attached to a circle the length of pSC101 DNA. Such lariats were consistently observed with this probe (Figure 4), but were not observed in a control reaction where the probe was replaced by P22 DNA that had been treated with λ exonuclease but not with terminal transferase (Figure 1). Since the one-ring heteroduplex between pDm4 and pSC101 DNAs also forms lariats with poly(dA)∼P22 DNA, we conclude that the poly(dA):poly(dT) joints survive the transformation and isolation steps.

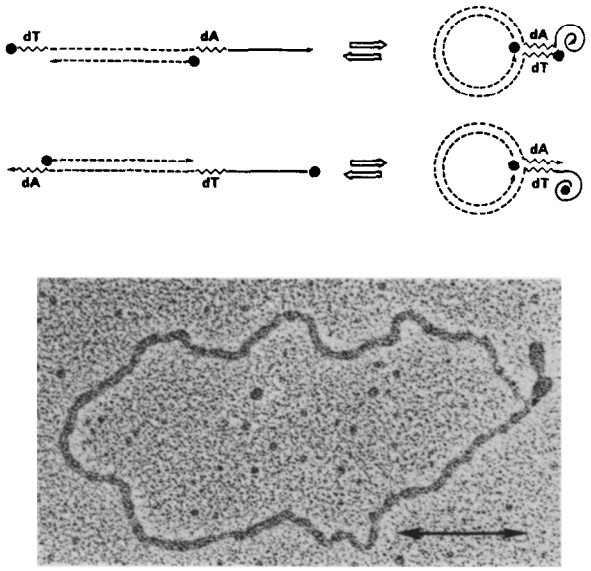

Figure 3. Heteroduplexes Between the Longest EcoR1 Fragment of pDm2 and EcoR1 cut pSC101

The symbols in the diagram have the same meaning as in Figure 1. Heteroduplexes were formed from EcoR1 cleaved pDm2 and pSC101 DNAs as described by Davis et al. (1971) except that the formamide was replaced by an equivalent volume of 0.2 M NaCl in TE, and renaturation effected by a 2 hr incubation at 45°C after a 2 min exposure to 65°C. The complementary linear heteroduplexes diagrammed at the left were observed in the electron microscope after spreading in 40% formamide. They consist of a double-stranded segment about the size of pSC101 DNA connected to a single-stranded tail about one half that length; a second single-stranded tail at the other end of the duplex segment could not be reliably visualized. When spread by the aqueous technique, circular double-stranded molecules the length of pSC101 DNA (9.2 ± 0.1 kb; N = 21) with an attached "blob" of collapsed single-stranded DNA were observed. An example is given in the electron micrograph, and its structure (see text) is depicted at the right in the diagram. The double-arrowed line represents 1.0 kb of duplex DNA.

Sequence Repetition is Not Observed within the pDm DNAs

The reassociation kinetics of randomly sheared pSC101 and pDm DNAs were determined as indicated in Figure 5. The data points fit the modified second order rate equation,

$$C/C_o = (1 + k\,C_o t)^{-n} \quad (1)$$

(sigmoid curve in Figure 5), or its equivalent,

$$(C_o/C)^{1/n} = 1 + k\,C_o t, \quad (2)$$

(inset to Figure 5), where C and C_o are the concentration of single-stranded DNA at time, t, and t = 0; k is the second-order rate constant, and n is an empirically determined constant that takes account of secondary reactions detected when the reassociation of random fragments is assayed with the single-strand specific S1 endonuclease (n = approximately 0.44; Morrow, 1974).

Table 2. Effect of DNA Polymerase on Closure of Heteroduplexes Described in Figure 3

	Relative Frequency of Circular Heteroduplexes Following Exposure to DNA Polymerase Plus:				
Experiment	dATP	dTTP	dGTP	dCTP	No dXTP
1	0.28	1.00	0.92	1.04	0.82
2	0.23	1.00	0.81	0.96	

The heteroduplexes of Figure 3 at a total DNA concentration of 8 μg/ml in 5 mM $MgCl_2$, 10 mM NaCl and 50 mM Tris-HCl(pH 7.5) were held at 65°C for 5 min, cooled quickly to 0°C, DNA polymerase and a nucleoside triphosphate were added to final concentrations of 4.4 μg/ml and 5×10^{-5} M, respectively, and the mixture kept at 15°C for 10 min when the reaction was stopped by addition of EDTA and NaCl to 11 mM and 110 mM, respectively. After incubation at 37°C for 2 hr to allow circularization, the mixture was aqueously spread (see Figure 3) and assayed for circular heteroduplexes. The assay consists of determining the ratio of circular heteroduplexes to SV40 DNA molecules which were added to the reaction mixtures as a concentration reference. The values in the table are these ratios normalized to that for the dTTP reaction.

Experiment 1: 600 SV40 molecules were counted for each reaction; number of circular heteroduplexes from dTTP reaction = 101.
Experiment 2: 67–70 SV40 molecules counted for each reaction except dCTP (33 counted); number of circular heteroduplexes from dTTP = 22.

When the reassociation of several nonrepetitious DNAs (for example, SV40, M13RF, and E. coli) was followed by this assay, the kinetics were observed to fit Equation (1) and yield values of k that are inversely proportional to the size of the genome (Morrow, 1974). This proportionality is also observed for the plasmid DNAs. Thus the ratios k_{pDm}/k_{pSC} (where the subscript indicates the plasmid DNA; see legend to Figure 5) and L_{pSC}/L_{pDm} (where L is the length of the plasmid DNA; Table 1) are, respectively, 0.77 and 0.76 for pDm1, 0.49 and 0.51 for pDm2, and 0.35 and 0.41 for pDm4. The agreement between the values for these two ratios and the close fit of the data points to Equation (1) or (2) indicate that most if not all sequences in each pDm DNA are represented only once per molecule.

The Inserted Sequences in pDm1 are Frequently Repeated in D. melanogaster; Those in pDm2 and 4 are Infrequently or Not Repeated

A large amount of D. melanogaster DNA was mixed with a small amount of ^{32}P-labeled plasmid DNA (w/w = 7×10^5), and the reassociation kinetics of the labeled DNA followed (Figure 6). The reassociation rate of pSC101 DNA is not affected by the D. melanogaster DNA, as is seen in Figure 6A by the agreement between the data points and the curve computed from Equation (1) using the value of k_{pSC} determined in Figure 5 (the reassociation of pSC101 DNA appears to occur at high $C_o t$ in Figure 6A because the $C_o t$ in the abscissa refers to the D. melanogaster DNA). By contrast, the reassociation rate of

that part of each pDm DNA which corresponds to the inserted segment is increased by the addition of D. melanogaster DNA. This effect is illustrated by the two-component curve given in Figure 6A for pDm1 DNA. One fourth of the labeled sequences reassociate rapidly along with the moderately repetitive sequences of D. melanogaster DNA, while the remaining three fourths exhibits the kinetics expected for pSC101 DNA. Similar two-component curves were obtained for pDm2 and pDm4 DNAs (not shown). In each case the fractions representing the inserted and pSC101 segments corresponded to the fast and slow reassociating fractions, with the slow fraction exhibiting the kinetics of pSC101 DNA.

Given the concentration of labeled pDm DNA, the fraction of this DNA that consists of the pSC101 sequences (Table 1) and the value of k_{pSC} (Figure 5), then Equation (1) can be used to compute the contribution of the pSC101 sequences to the total reassociated DNA for each data point. The values of C^*/C_o^* for the total labeled DNA (ordinate to Figure 6A) can then be transformed to C^{Dm}/C_o^{Dm}, where C^{Dm} represents the single-stranded concentration of labeled D. melanogaster sequences from the plasmid. A computer program was written to effect this transformation for each data point, and the results are given in Figure 6B, along with curves that represent the best fit of Equation (1) to the data.

The good fit between these curves and the transformed data points suggests that the inserted sequences in a given pDm form a homogeneous class with respect to their frequency of occurrence within the D. melanogaster genome. This frequency, α, can then be calculated from the following equation which is derived in Footnote (a) of Table 3:

$$\alpha = (k'_{pDm}/k_{pDm})(L_D/L_{pDm}) \quad (3)$$

where k_{pDm} and L_{pDm} have their previously assigned meanings; k'_{pDm} is the value of k obtained when Equation (1) is fitted to the data in Figure 6B; and L_D is the length of DNA in the D. melanogaster genome (165,000 kb; Rudkin, 1972). The values of α for each pDm DNA are given in Table 3.

The Inserted Sequences in pDm1 are Located at Many Different Sites in the D. melanogaster Chromosomes: Those in pDm2 and 4 are at Single Sites

^{3}H-labeled RNA copies (cRNA) of the sequences in each pDm DNA were transcribed with E. coli RNA polymerase and hybridized to D. melanogaster polytene chromosomes in situ as described in the legend to Figure 7. Autoradiographs obtained with cRNA from either pDm2 or pDm4 exhibited silver grains over a single chromomeric region (Figure 7; Table 3) whether the exposure was for 13, 41, or 76 days. By contrast, cRNA from pDm1 hybridized to the β-heterochromatin in the chromocenter, and to 15 different chromomeric regions (Table 3). As expected, cRNA from pSC101 exhibited no in situ hybridization, even after 41 days of exposure.

Discussion

One Hybrid–One Chromosomal Segment

The pDm DNAs offer a means of obtaining reliable sequence maps of D. melanogaster chromosomes provided that the inserted DNA represents a single continuous segment of the chromosomal DNA. The poly(dA):poly(dT) method for joining DNA molecules from two different sources is designed to satisfy this provision since linkage between molecules from the same source does not occur (Jackson et al., 1972; Lobban and Kaiser, 1973). Circular hybrids formed by this method should therefore consist of an equal number of pSC101 and chromosomal segments that alternate around the circle. The structures of the pDm/pSC101 heteroduplexes indicate that this number is one for each of the pDm DNAs.

Other structural characteristics of the hybrid plasmids confirm that the D. melanogaster DNA in each is derived from a single chromosomal segment inserted by this method. We thus have demonstrated that insertion occurred at the EcoR1 site in the pSC101 DNA, and that the two DNAs are linked by poly(dA):poly(dT) duplexes.

The single insertion characteristic of the poly(dA):poly(dT) method is not inherent to the EcoR1–ligase joining method (Mertz and Davis, 1972). In this method, the interacting single-stranded tails of the different DNA molecules are identical, and ad hoc provisions are required to insure that the inserted foreign DNA does not consist of multiple segments linked to each other. Another difference is that the segments to be joined by the poly(dA):poly(dT) method can be created by virtually any mechanism from shear breakage to cleavage by specific endonucleases, whereas the EcoR1–ligase method requires that all segments must be formed by EcoR1 cleavage at both ends. We chose shear breakage because we wish to build a library of randomly selected segments from the D. melanogaster genome.

The Dm2, 4 Class of Chromosomal Segments

In considering the properties of these segments, we shall refer to them as Dm1, Dm2, etc., where the number refers to the respective pDm plasmid. Though the present library is small, it is convenient to divide it into two classes. Dm2 and Dm4 belong to one class defined by a low repetition frequency,

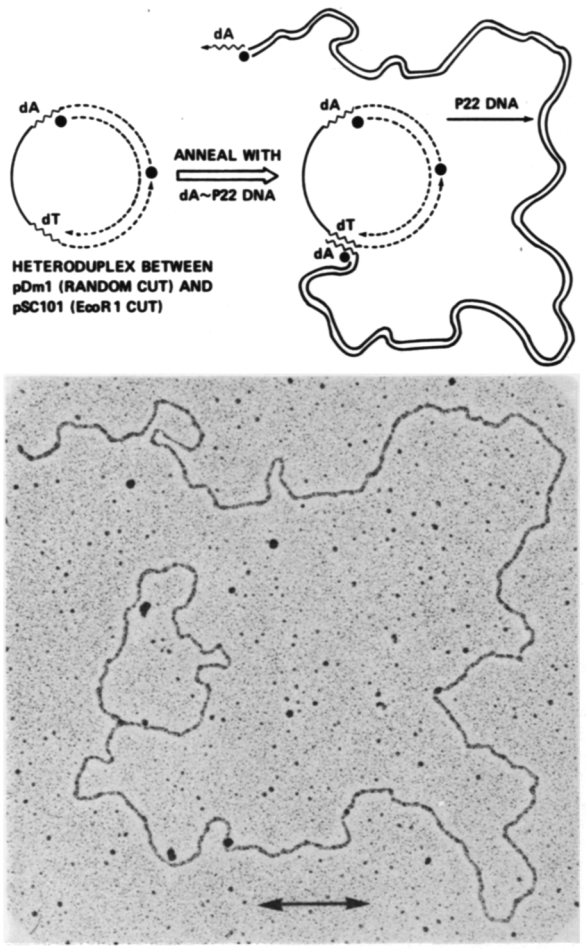

Figure 4. Attachment of Poly(dA)~P22 DNA to Heteroduplexes Formed Between pDm1 and pSC101 DNAs

The heteroduplex depicted at the left in the diagram was formed from EcoR1 cut pSC101 DNA (Figure 1) and randomly nicked circular pDm1 DNA by the method indicated in Figure 2. Formamide was removed by dialysis against 0.1 M NaCl in TE, and this DNA (0.8 μg/ml) allowed to interact with poly(dA)~P22 DNA (4.5 μg/ml) in this solvent for 2 hr at 45°C after a 2 min exposure to 65°C. The molecule in the electron micrograph was observed after aqueous spreading; the single-stranded region (see diagram) is collapsed under this condition. Lengths of the circular and tail portions of such molecules (N = 7) are 8.9 (±0.3) kb and 39.0 (±0.8) kb, and correspond to the lengths of pSC101 (Table 1) and P22 (Rhoades, MacHattie, and Thomas, 1968) DNAs. The double-rowed line represents 2.0 kb of duplex DNA.

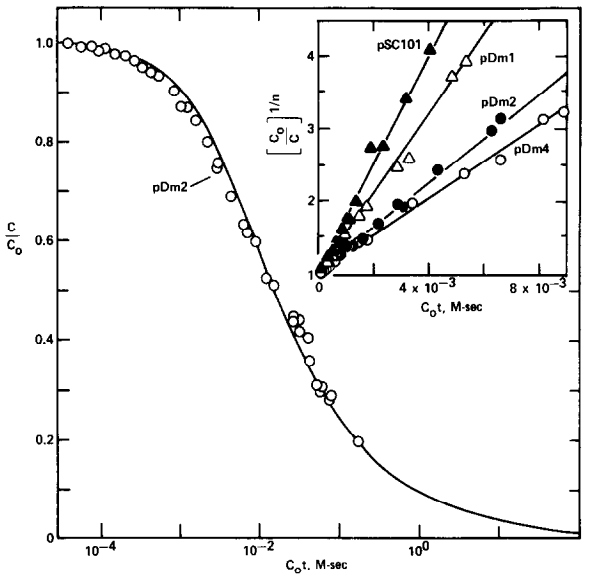

Figure 5. Reassociation Kinetics of Plasmid DNAs

Plasmid DNAs were labeled in vitro with ^{32}P to 1–2 × 10^8 cpm/μg (Schachat and Hogness, 1973) and mixed with a 100 fold excess of unlabeled plasmid DNA to obtain a C_o of about 2 × 10^{-7} M in the reassociation reaction. Shearing, denaturation and reassociation (assayed with S1 endonuclease) were carried out in the presence of calf thymus DNA (250 μg/ml) as described by Schachat and Hogness (1973). Reassociation occurred at 65°C in 0.5 M NaCl in TE. The sigmoid curve and the straight lines in the inset are the least squares best fit of Equations (1) and (2) [n = 0.42, the mean value for 0.5 M NaCl; Morrow (1974)] to the data, and yield k_{pSC} = 539, k_{pDm1} = 415, k_{pDm2} = 264, and k_{pDm4} = 189/M · sec.

α, and localization to a single chromomeric region. Dm1 must be placed in another class because α is greater by more than an order of magnitude and its sequences are found within the chromocentric β-heterochromatin as well as in at least 15 different chromomeric regions.

Neither segment in the Dm2, 4 class can consist of a tandem array of 2 or more repeating units that encompass the segment. These sequence arrangements would significantly alter the kinetics of reassociation for the pDm2 and 4 DNAs from that observed in Figure 5. High tandem repeat numbers would yield curves that were clearly biphasic, and the lower repeat numbers would cause the k_{pDm} of best fit to be at least 40% greater than that expected from its contour length, L_{pDm}. The observed curves do not exhibit a significant biphasic character and yield values for k_{pDm2} and k_{pDm4} that are, respectively, 4 and 15% less than that expected from the respective L_{pDm}s.

However, the α values of Dm2 and 4 suggest that the entire segment of each may be repeated within the genome and the results of in situ hybridization indicate that such a repetition would be highly localized. The primary determinant of the error in α is the ratio, k'_{pDm}/k_{pDm}, in Equation (3). Both rate constants were determined with the same preparation of labeled pDm DNA under very similar conditions (Figures 5 and 6). We therefore expect a cancelling out in the ratio of the effects of such factors as the length distribution of the reassociating fragments and their base composition and sequence. The accuracy of α ought then to be largely determined by the experimental reproducibility in determining the k'_{pDm}/k_{pDm} ratio. On this basis we conservatively estimate that the values of α for Dm2 and 4 are accu-

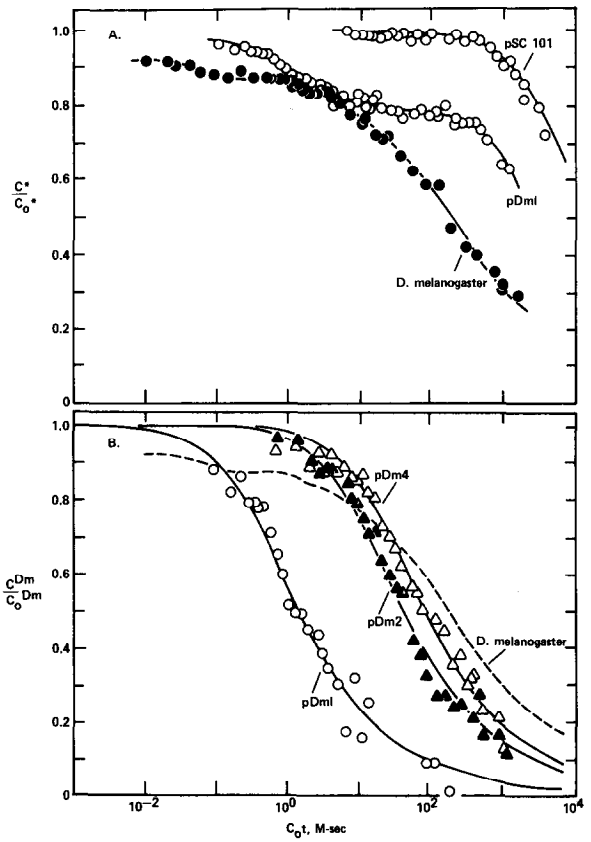

Figure 6. Reassociation Kinetics of Plasmid DNAs in the Presence of D. melanogaster DNA

Reassociation was carried out as described in Figure 5 except that the unlabeled pDm DNA and calf thymus DNA were omitted and D. melanogaster DNA was added to a concentration (approximately 1×10^{-2} M) that was 7×10^5 times that of the labeled plasmid DNA. C_o in the abscissa refers to the D. melanogaster DNA.

(A) The ordinate, C^*/C_o^*, refers to the labeled plasmid DNA for the pSC101 and pDm1 curves, and to the labeled D. melanogaster DNA for the third curve (plasmid DNA absent).

(B) The ordinate, C^{Dm}/C_o^{Dm}, represents the labeled D. melanogaster sequences in the pDm DNA (see text). The curves are the least squares best fit of Equation (1) to the data points in this figure [that is, $C^{Dm}/C_o^{Dm} = (1 + k'_{pDm}C_ot)^{-n}$, where n = 0.42, and C_o refers to D. melanogaster DNA], and correspond to: $k'_{pDm1} = 2.74$; $k'_{pDm2} = 0.101$; and $k'_{pDm4} = 0.049/M \cdot sec$. The data points for pDm1 are from one of the two experiments used to provide the data for the pDm1 curve in Panel A.

rate to within a factor of two, or better. Hence, we are reasonably confident that the Dm2 segment is at least duplicated within its chromomeric region, but we must reserve judgement regarding the Dm4 segment.

Figure 6B indicates that the Dm2 and 4 segments do not contain a short period interspersion of repetitive sequences of the dimensions observed in Xenopus laevis (Davidson et al., 1973) and sea urchins (Graham et al., 1974). The average interspersion pattern in these two animals consists of 0.3 kb stretches of repetitive sequences that alternate with 0.7–1.0 kb stretches of nonrepetitive sequences. Since the repetitive sequences constitute about one fourth of this interspersion pattern and exhibit repetition frequencies more than an order of magnitude greater than the α values for Dm2 and 4, they should cause the data points at the lower C_ot values to fall well below the fitted curves in Figure 6B.

Although such a marked fall off of the data points is not observed, it should be emphasized that an interspersion pattern in which the stretch of repetitive sequences is considerably shorter, or in which the interspersion period is considerably longer, would not be reliably detected by these reassociation kinetics. For example, the fall off of the early data points from the pDm4 curve would indicate a more rapidly reassociating minority fraction representing about one tenth of the segment were it not for the fact that the deviation of these points from the curve is at the borderline of significance.

Table 3. Frequency of Repetition, α, and Chromosomal Location of D. melanogaster Sequences in pDm DNAs

Plasmid	α[a]	Location[b]
pDm1	90.0	Chromocenter (β-heterochromatin)
		+ 15 chromomeric regions;
		X: 2A, 18C, 19E
		2L: 27E, 32B
		2R: 43D, 56F, 59B, 59F
		3L: 63A
		3R: 84A, 87E, 88B, 97B, 99D-E
pDm2	3.5	1 chromomeric region; 3R: 84D
pDm4	2.0	1 chromomeric region; 3L: 62E

[a] The definition of α is given by $\alpha = FL_D/fL_{pDm}$, or $\alpha = (F/f)(L_D/L_{pDm})$, where F and f are, respectively, the fraction of L_D and L_{pDm} represented by the sequences in the inserted DNA. The second order rate constant for the inserted DNA, k_{Dm} is given by $k_{Dm} = k_{pDm}/f$, or by $k_{Dm} = k'_{pDm}/(F + fR)$, where R is the ratio of pDm to D. melanogaster DNAs used in Figure 6 (that is, 1.4×10^{-6}). Hence, $(F + fR)/f = k'_{pDm}/k_{pDm}$. Since $F >> fR$ under conditions used here (for example, F/fR = 53 for an inserted segment of 3 kb represented only once per D. melanogaster genome), then k'_{pDm}/k_{pDm} can be substituted for F/f in the equation defining α to obtain Equation (3), which was used to calculate α.

[b] See Lindsley and Grell (1968) for numeration of these chromomeric regions, which contain from 4–17 bands. The autoradiographic resolution is to within a few bands, and hence in many cases better than to one of the indicated regions.

The Dm1 Chromosomal Segment

The sequences in the Dm1 segment contrast in almost every characteristic with those in the preceding class. Repeated about 90 times, they represent some 260 kb, or 0.16% of the genome, and on this ground alone are unlikely to be confined to a single chromomeric unit. Indeed, the in situ hybridization

results indicate that these sequences are interspersed at long intervals within the chromosomal DNA (Table 3).

There are two extreme interpretations of these results. All sequences within the Dm1 segment could be located in each of the 16 chromosomal regions. Or these sequences could be divided into 15 subsets, such that one chromosomal region contains all 15 (that is, the chromosomal segment isolated in pDm1), and the other regions each contain a different one of the subsets. The fact that the data points in Figure 6B exhibit a good fit to Equation (1), and hence do not indicate that the Dm1 sequences are heterogeneous with respect to α, is most simply explained by the first of these interpretations.

We cannot at this stage eliminate the other extreme interpretation, or variants of it, if the sequences in the postulated subsets exhibit α values that are clustered about 90 repeats per genome. The simplest method for distinguishing among these interpretations is to divide the Dm1 segment into subsegments, and to determine for each isolated subsegment the chromosomal location of its sequences by in situ hybridization.

An analysis of the function of the Dm1 sequence must await this kind of sequence mapping. At this stage we wish to point out only that the wide but specific distribution of these sequences in the D. melanogaster chromosomes can be fit to many regulatory models for the eucaryotic cistron of which the prototype is that proposed by Britten and Davidson (1969). In this regard, we note that the pattern of silver grains over the 56F regions that is given by the Dm1 sequences (Table 3) is indistinguishable from that given by ^{3}H-labeled 5S ribosomal RNA

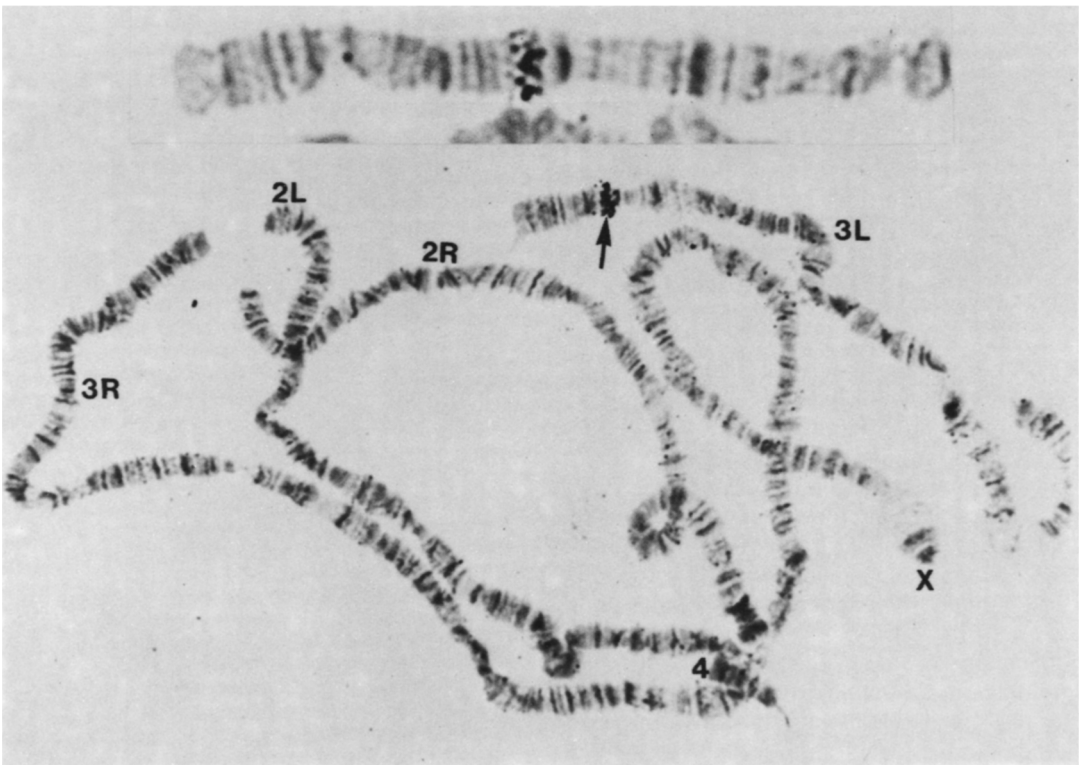

Figure 7. In situ Hybridization of pDm4 cRNA

^{3}H-cRNA was synthesized from 2.5 μg of pDm4 DNA in 100 μl of 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 40 mM Tris-HCl (pH 8.0) containing 40 nmoles of GTP and 50 μCi each of ^{3}H-labeled ATP, CTP, and UTP (34, 31, and 22 Ci/m mole respectively). Five μg E. coli RNA polymerase was added and the mixture incubated at 37°C. After 90 min, 400 μl of 100 μg/ml yeast RNA in 40 mM Tris-HCl (pH 7.8) and 10 μg DNAase I were added. Incubation for 20 min at 25°C was followed by phenol and ether extraction. cRNA was separated from nucleoside triphosphates by Sephadex G-50 chromotography. Cytological squashes of salivary gland chromosomes were prepared from *gt/gt*^{X11} larvae (Lindsley and Grell, 1968) by standard methods. In situ hybridization was performed as described (Pardue et al., 1973), except that cRNA was dissolved in 0.33 M NaCl instead of 2 × SSC. The hybridization mixture combined 2 mg/ml yeast RNA and 1.0 × 10^7 to 2.5 × 10^7 cpm/ml of ^{3}H-cRNA and was heated at 85°C for 5 min immediately before use. Twenty μl of this mixture were used for each slide. Exposure times were 13 and 41 days for the upper (2200x) and lower (550x) micrographs. The silver grains are clustered over the 62E region in 3L (Lindsley and Grell, 1968), that is, the left arm of the third chromosome. This arm has broken away from the chromocenter, which is located at the lower right of the figure.

[M. L. Pardue, personal communication; also see Wimber and Steffensen (1970) and Grigliatti et al. (1973)]. Perhaps certain of the sequences in Dm1 are involved in the regulation of the expression of the 5S genes.

Concluding Remarks

The Dm2, 4 class of segments may represent the majority of the chromomeric DNA. Thus 4 of 6 additional Dm segments that have recently been isolated fit the definition of this class, in that their sequences exhibit small values of α (0.9–2.0) and are restricted to single chromomeric regions (Glover, White, Finnegan, and Hogness, in preparation). The apparent high frequency of this class and its characteristics provide a strong argument against the generality of the tandem repetition models for the chromomeric unit (Callan, 1967; Thomas, 1970; Lee and Thomas, 1973).

These models explain the anomalous DNA to cistron ratio for chromomeric units (see Introduction) by the postulate that most of the chromomeric DNA consists of tandemly repeated sequences, each coding for the same polypeptide. The Dm2 and Dm4 segments may be part of such an array, but if so, the repeating unit is at least as large as the segment, or 8.7 and 12.6 kb, respectively. The same is true for the 4 new Dm segments in this class, which have lengths between 15 and 19 kb. The molecular weight of polypeptides coded by such segments would range from 320,000 to 700,000 daltons. While structural genes of this size are not impossible, it is highly unlikely that all six segments would derive from such large genes.

How then can one explain the data for Dm2, where the evidence for repetition is reasonably convincing? Perhaps Dm2 represents a duplication (or triplication, etc.) of an entire chromomeric unit. Lefevre and Green (1972) have described a duplication in the 3C region that appears to be of this kind, and Lefevre (1973) has noted other regions where such duplications probably occur, each containing a pair of adjacent dark-staining doublet bands that characterizes the duplication in 3C. We note that the Dm2 segment is located in a region containing such a pair of doublets (84D 1–2 and 3–4), and that the silver grains are centered over this part of the 84D region. If this suggestion is correct, then in an expanding library of segments, the Dm2, 4 class should divide into a majority subclass of segments that for the most part consist of nonrepetitive sequences, and a minority subclass of segments that are locally repeated a few times.

In any case, the above suggestion does not provide a solution to the structure of the chromomeric unit. The viability of additional models for this unit (for example, see Crick, 1971; Paul, 1972; Prescott and Murti, 1973; O'Brien, 1973) can be tested by detailed mapping of sequences within the segments —particularly with respect to undetected repetitive sequences that may represent a small fraction of the segment, and sequences present in hnRNA and mRNA of different cell types.

We conclude with a word of caution regarding these future mapping expeditions. Although the data indicate that the Dm segments derive from single chromosomal segments, we do not know that they have not been altered (by deletion, for example) during transformation and amplification. What information we do have does not indicate any instability. Thus two 17 generation subcultures of each of the three transformants yielded pDm molecules with the same size distributions, each of which indicated the molecules to be as homogeneous in length as pSC101 DNA. However, this test would not detect minute deletions and is not relevant to alterations which may take place early in transformation. In this regard it should be realized that we used a recombination-proficient (*rec*$^+$) recipient strain for these transformations. Clearly a recombination-deficient recipient would be preferable, and indeed the 6 new hybrids referred to above were isolated by transformation of *recA*$^-$ strains which are absolutely defective for host recombination.

Experimental Procedures

Bacteria

E. coli K12 strain C600 (Appleyard, 1954) lacking modification (*hsm*$^-$) and restriction (*hsr*$^-$) functions and carrying a deletion of the *gal* operon was obtained from R. W. Davis and is designated h303; C600 (pSC101) was obtained from S. N. Cohen (Cohen et al., 1973).

DNAs

D. melanogaster (Oregon R) DNA was isolated from 0–15 hr embryonic nuclei according to the method of Schachat and Hogness (1973) with minor modifications. Closed circular pSC101 and pDm DNAs were isolated from lysates (Clewell and Helinski, 1969) of C600 (pSC101) and h303 (pDm) by centrifugation in ethidium bromide/CsCl gradients (Radloff, Bauer, and Vinograd, 1968). Ethidium bromide and contaminating RNA were removed by passage through a column (diameter = 1 cm) with a 2 cm layer of Dowex AG50 on top of 13 cm of Biogel A15M. The column was equilibrated and the DNA eluted with 0.5 M NaCl in TE (1 mM EDTA, 10 mM Tris–HCl, pH 8.2). All DNA solutions were dialyzed against TE prior to storage at 4°C.

Enzymes

EcoR1 endonuclease (Green et al., 1974), λ exonuclease (Little, Lehman, and Kaiser, 1967), calf thymus terminal transferase (Kato et al., 1967) and E. coli RNA polymerase (Berg, Barrett, and Chamberlin, 1971) were prepared according to the indicated references and generously provided by D. M. Glover and R. L. White, by P. E. Lobban, by R. L. Ratliff and J. L. Hanners, and by M. Chamberlin and W. Wickner, respectively. DNAase I and Aspergillus oryzae S1 endonuclease were obtained from Worthington Biochem. Corp. and Seikagaku Kogyo Co., respectively.

Electron Microscopy

DNA was spread by the aqueous or 40% formamide methods (Davis, Simon, and Davidson, 1971) as indicated in the Figure legends, and contour lengths were measured (Schachat and Hogness, 1973) relative to reference DNAs present on the same grid.

Acknowledgment

We thank Mary Lou Pardue for introducing us to the technique of in situ hybridization and for scoring many of the autoradiographs. We also thank Louise Prestidge for her assistance in the electron microscopic characterization of the hybrids. This work was supported by research grants from the National Science Foundation and the National Institutes of Health. P. C. W. was a Dernham fellow and more recently, a fellow of the American Cancer Society. D. J. F. is a Jane Coffin Child fellow. J. E. D. was a Helen Hay Whitney fellow.

Received August 30, 1974

References

Appleyard, R. K. (1954). Genetics *39*, 440–452.

Beermann, W. (1972). In Results and Problems in Cell Differentiation, *4*, W. Beermann, ed. (New York: Springer-Verlag), pp. 1–33.

Berg, D., Barrett, K., and Chamberlin, M. (1971). In Methods in Enzymology, *21D*, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 506–519.

Bishop, J. O. (1974). Cell *2*, 81.

Britten, R. J., and Davidson, E. H. (1969). Science *165*, 349–357.

Callan, H. G. (1967). J. Cell Sci. *2*, 1–7.

Champoux, J. J., and Hogness, D. S. (1972). J. Mol. Biol. *71*, 383–405.

Clewell, D. D., and Helinski, D. R. (1969). Proc. Nat. Acad. Sci. USA *62*, 1159–1166.

Cohen, S. N., Chang, A. C. Y., Boyer, H. W., and Helling, R. B. (1973). Proc. Nat. Acad. Sci. USA *70*, 3240–3244.

Crick, F. H. C. (1971). Nature *234*, 25–27.

Davidson, N., and Szybalski, W. (1971). In The Bacteriophage Lambda, A. D. Hershey, ed. (Cold Spring Harbor, New York: Cold Spring Harbor Labs), pp. 45–82.

Davidson, E. H., Hough, B. R., Amenson, C. S., and Britten, R. J. (1973). J. Mol. Biol. *77*, 1–23.

Davis, R. W., Simon, M., and Davidson, N. (1971). In Methods in Enzymology, *21D*, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 413–428.

Green, P. J., Bellach, M. C., Goodman, H. M., and Boyer, H. W. (1974). In Methods in Molecular Biology, *6*, R. Wickner, ed. (New York: Marcel Dekker), in press.

Graham, D. E., Neufeld, B. R., Davidson, E. H., and Britten, R. J. (1974). Cell *1*, 127–137.

Grigliatti, T. A., White, B. N., Tener, G. M., Kaufman, T. C., Holden, J. J., and Suzuki, D. T. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 461–474.

Hochman, B. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 581–589.

Hogness, D. S., and Simmons, J. R. (1964). J. Mol. Biol. *9*, 411–438.

Jackson, D., Symons, R., and Berg, P. (1972). Proc. Nat. Acad. Sci. USA *69*, 2904–2909.

Judd, B. H., Shen, M. W., and Kaufman, T. C. (1972). Genetics *71*, 139–156.

Kato, K.-L., Goncalves, J. M., Houts, G. E., and Bollum, F. J. (1967). J. Biol. Chem. *242*, 2780–2789.

Laird, C. D. (1973). Ann. Rev. Genetics *7*, 177–204.

Lee, C. S., and Thomas, C. A., Jr. (1973). J. Mol. Biol. *77*, 25–42.

Lefevre, G., Jr. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 591–599.

Lefevre, G., Jr., and Green, M. M. (1972). Chromosoma *36*, 391–412.

Lennox, E. S. (1955). Virology *1*, 190–206.

Lindsley, D. L., and Grell, E. H. (1968). In Genetic Variations of Drosophila melanogaster (Washington, D.C.: Carnegie Institution).

Little, J. W., Lehman, I. R., and Kaiser, A. D. (1967). J. Biol. Chem. *242*, 672–678.

Lobban, P. E., and Kaiser, A. D. (1973). J. Mol. Biol. *78*, 453–471.

Mandel, M., and Higa, A. (1970). J. Mol. Biol. *53*, 159–162.

Marvin, D. A., and Hohn, B. (1969). Bacteriol. Rev. *33*, 172–193.

Mertz, J. E., and Davis, R. W. (1972). Proc. Nat. Acad. Sci. USA *69*, 3370–3374.

Morrow, J. F. (1974). Ph.D. Thesis, Stanford University, Stanford, California.

Morrow, J. F., and Berg, P. (1972). Proc. Nat. Acad. Sci. USA *69*, 3365–3369.

Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Goodman, H. M., and Helling, R. B. (1974). Proc. Nat. Acad. Sci. USA *71*, 1743–1747.

O'Brien, S. J. (1973). Nature New Biol. *242*, 52–54.

Pardue, M. L., Brown, D. D., and Birnsteil, M. L. (1973). Chromosoma *42*, 191–203.

Paul, J. (1972). Nature *238*, 444–446.

Peacock, W. J., Brutlag, D., Goldring, E., Appels, R., Hinton, C. W., and Lindsley, D. L. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 405–416.

Prescott, D. M., and Murti, K. G. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 609–618.

Radloff, R., Bauer, W., and Vinograd, J. (1968). Proc. Nat. Acad. Sci. USA *59*, 838–845.

Rhoades, M., MacHattie, L. A., and Thomas, C. A., Jr. (1968). J. Mol. Biol. *37*, 21–40.

Rudkin, G. T. (1972). In Results and Problems in Cell Differentiation, *4*, W. Beerman, ed. (New York: Springer-Verlag), pp. 59–85.

Schachat, F. H., and Hogness, D. S. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 371–381.

Thomas, C. A., Jr. (1970). In The Neurosciences: Second Study Program, F. O. Schmitt, ed. (New York: The Rockefeller University Press), pp. 973–998.

Wimber, D. E., and Steffensen, D. M. (1970). Science *170*, 639.