Cell, Vol. 5, 149–157, June 1975, Copyright © 1975 by MIT

Characterization of Six Cloned DNAs from Drosophila melanogaster, Including One that Contains the Genes for rRNA

David M. Glover, Raymond L. White, David J. Finnegan, and David S. Hogness*
Department of Biochemistry
Stanford University School of Medicine
Stanford, California 94305

Summary

pDm plasmids were constructed from D. melanogaster and pSC101 DNAs by a modification of the EcoR1-ligase method which insured that each hybrid molecule contained a single segment of D. melanogaster chromosomal DNA (Dm segment). The sequences in the Dm segments of six cloned pDm DNAs were mapped within the D. melanogaster polytene chromosomes by in situ hybridization, and their repetition frequencies within the Dm segment and within the genome were determined. Four of these segments consist of sequences that are confined to single chromomeric regions in the polytene chromosomes and exhibit little or no repetition. The characteristics of this group, and also two of three Dm segments analyzed earlier (Wensink et al., 1974), are inconsistent with tandem repetition models of the chromomere. By contrast, the other two Dm segments contain moderately repetitive sequences that are located in the heterochromatin. One of these appears to be a segment of the Y chromosome in which about half the sequences are nonrepetitive and half are repeated about 33 times per genome, though they are not repeated within the segment. The second contains the DNA coding for 18 and 28S rRNA.

Introduction

The chromomeric units in the chromosomal DNA of Drosophila melanogaster are recognized by the bands they form in polytene chromosomes, and are individually defined as that segment of the DNA which contains the sequence of base pairs found in a band plus an adjacent interband (Hogness et al., 1974). Cytogenetic analyses suggest that each of the approximately 5000 chromomeric units in the genome contains only one cistron (Judd, Shen, and Kaufman, 1972; Hochman, 1973), even though the mean length of these units is 26 kilobase pairs (kb) —that is, enough DNA to code for about two dozen typical polypeptides (Hogness et al., 1974; Laird, 1973). [Abbreviations: kb (kilo bases) equals 1000 bases or base pairs in single- or double-stranded nucleic acids, respectively.]

Our present ignorance of the arrangement of genetic information in eucaryotic chromosomes allows the coexistence of a variety of models to explain this low cistron to DNA ratio (Beerman, 1962; Bishop, 1974). These models can, however, be distinguished according to the topography that each imposes on the repetitive and nonrepetitive sequences, and on the transcribed and translated sequences. We therefore have initiated a series of experiments designed to map these sequences within and among the chromomeric units. In the first report on this series (Wensink et al., 1974) three different segments of D. melanogaster DNA (abbreviated by Dm segments) were isolated and their sequences analyzed for repetition and mapped within the polytene chromosomes. They were isolated by constructing circular hybrid DNAs consisting of a single Dm segment linked by poly(dA) • poly(dT) joints to a single molecule of the tetracycline resistance plasmid, pSC101, and cloning such hybrid plasmids (abbreviated by pDm) by introducing them into Escherichia coli and selecting for tetracycline-resistant colonies.

Here we report on the characterization of the sequences in six additional Dm segments, also isolated by cloning pDm plasmids. In this case the Dm segments were inserted into pSC101 by the EcoR1-ligase method (Mertz and Davis, 1972; Cohen et al., 1973), modified to guarantee that each pDm plasmid contains only one Dm segment. Four of the six segments consist of sequences that exhibit little or no repetition and are confined to a single chromomeric region. These properties also characterize two segments in the previous set (Wensink et al., 1974), and are inconsistent with tandem repetition models for the chromomeric unit. The other two segments analyzed here contain moderately repetitive sequences that are located in the heterochromatin. One of these contains the DNA complementary to 18S and 28S rRNAs.

Results

The pDm Molecules Contain a Single Dm Segment

The EcoR1-ligase method for linking DNA fragments is indiscriminate in that any two of the identical termini created by the EcoR1 endonuclease can be covalently joined with E. coli ligase (Mertz and Davis, 1972; Cohen et al., 1973). A mixed population of circular molecules with the general formula, $(Dm)_i(pSC)_j$, therefore is expected from the ligation of a mixture of EcoR1-cut pSC101 DNA (abbreviated by pSC) and the Dm segments created by cleavage with this enzyme. To insure that hybrids with i or $j > 1$ are excluded during the isolation of the desired $(Dm)_1(pSC)_1$ hybrid, we have introduced the following series of molecular sizing steps.

*To whom reprint requests should be sent.

Table 1. Lengths of the pDm Plasmids and the EcoR1 Fragments Obtained from Their Dm Segments

Plasmid	L_{pDm}[a] (kb)	L_{Dm}[b] (kb)	EcoR1 Fragments[c] (kb)
pDm101	24.6 (±0.2)	15.4	**9, 3.8** **1.5, 1.4** Σ = 15.9
pDm102	27.7 (±0.3)	18.5	11.8, 6.1 Σ = 17.9
pDm103	26.3 (±0.3)	17.1	17.0
pDm105	26.8 (±0.3)	17.6	4.9, 3.3 **2.9, 2.9** **2.6, 1.4** Σ = 18.0
pDm106	24.5 (±0.4)	15.3	7.9, 6.1 1.3 Σ = 15.3
pDm107	27.2 (±0.2)	18.0	8.6, **4.9** **4.3** Σ = 17.8

[a]The mean length (L_{pDm}) and its standard error (in parentheses) are given for each plasmid (N = 17 ± 4).
[b]$L_{Dm} = L_{pDm} - L_{pSC}$, where L_{pSC} is 9.2 kb, the length of pSC101 DNA (Wensink et al., 1974).
[c]The lengths of the EcoR1 fragments were determined by electron microscopy, except for the values given in bold face print. These were determined by agarose gel electrophoresis (Green et al., 1974), using the EcoR1 fragments of phage λ DNA (Thomas and Davis, 1975) and the EcoR1 and Hpa I fragments of SV40 DNA (Nathens and Smith, 1975) as length standards.

D. melanogaster DNA was partially cleaved with EcoR1 and then fractionated by agarose gel electrophoresis to obtain a narrow size class of Dm segments with mean length and standard deviation of 14.2 ± 2.3 kb (N = 126). These molecules were incubated with EcoR1-cut pSC101 DNA in the presence of ligase, and the reaction products fractionated by zone sedimentation in a sucrose gradient to eliminate circular hybrids larger than $(Dm)_1(pSC)_1$ (see Experimental Procedures). The DNA in the combined gradient fractions was then used to transform tetracycline-sensitive E. coli HB101 to resistance according to previously described procedures (Wensink et al., 1974).

A trial screening indicated that only a small fraction (1/40) of the transformants contained pDm plasmids, the remainder containing pSC101. This low frequency is due, at least in part, to the stringent selection against the $(Dm)_2(pSC)_1$ or larger hybrids that was employed when combining the gradient fractions (Experimental Procedures). In order to increase the proportion of hybrid molecules, plasmids were isolated from a mass culture inoculated with bacteria from approximately 800 colonies, and this DNA was fractionated by gel electrophoresis.

The plasmid DNAs isolated from the mass culture distributed into two length classes corresponding to pSC101 monomers and the $(Dm)_1(pSC)_1$ hybrids. As expected, the hybrid class constituted only a few percent of the total and had a length distribution of 22.6 ± 2.7 kb (N = 37). The hybrid molecules, separated from pSC101 monomers by gel electrophoresis, were used to obtain individual transformed clones. Seven of these clones were selected and their plasmid DNAs characterized with respect to the properties given in Table 1. Only six of the pDm plasmids are listed in the table because two (pDm 101 and 104) are identical with respect to all properties discussed in this paper. We presume that both originate from the same one of the approximately 20 hybrid clones in the original population of 800 transformants.

The length distributions of these plasmids indicate that each is homogenous, and that these lengths correspond to the desired $(Dm)_1(pSC)_1$ hybrids. This 1:1 structure is also indicated by the population of fragments obtained by complete EcoR1 cleavage. Each plasmid yielded a pSC101-size fragment (9.2 kb; Wensink et al., 1974), together with the fragments listed in Table 1, which are presumed to derive from the Dm segment. Whereas five of the pDm DNAs yielded one mole of the pSC101-size fragment per mole of the other fragments, one (pDm101) produced 2 moles of 9 kb fragments per mole of the others. As is indicated in Table 1, we have assigned one of these to the Dm101 segment. The conclusion that pDm101 has a $(Dm)_1(pSC)_1$ rather than a $(Dm)_1(pSC)_2$ structure is based on the following two arguments.

—The reassociation kinetics of pDm101 DNA in the presence of total D. melanogaster DNA (reported in a subsequent section) indicate that approximately 60% of this hybrid consists of D. melanogaster sequences. The expected percentages are 63% and 25% for the $(Dm)_1(pSC)_1$ and $(Dm)_1(pSC)_2$ structures, respectively.

—Two pSC101 segments would occupy 18.4 kb of the 24.6 kb in pDm101, and leave only 6.2 kb for the Dm segment, which is more than 3 standard deviations less than the mean length of the Dm segments used to construct the hybrids.

The Buoyant Densities of the Dm Segment Correspond to that for D. melanogaster DNA

The buoyant densities of pDm101, 102, 103, 105, 106, and 107 DNAs in CsCl are, respectively, 1.7045, 1.7065, 1.7047, 1.7029, 1.7069, and 1.7058 g · cm^{-3}. Given that the density of pSC101 DNA is 1.7104 g · cm^{-3}, and that the relative lengths of the Dm and pSC101 segments in each hybrid are known (Table 1), one can calculate the buoyant densities of each Dm segment (Wensink et al., 1974). These densities are: 1.701—Dm101, 1.705—Dm102, 1.702—Dm103, 1.699—Dm105, 1.706—Dm106, and 1.703—Dm107. All values fall within the density profile of the main band of D. melanogaster DNA (Schachat and Hogness, 1973).

The Dm Segments Divide into Two Groups According to the Repetition Frequencies of Their Sequences

The reassociation kinetics of randomly sheared pDm DNAs have been determined in the absence and presence of total D. melanogaster DNA to test for sequence repetition within the Dm segment and within the D. melanogaster genome, respectively. The self driven reassociation kinetics were determined as indicated in Figure 1. The data points for all pDm DNAs except pDm103 fit the modified second order rate equation

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

as is exemplified by the results given for pDm105

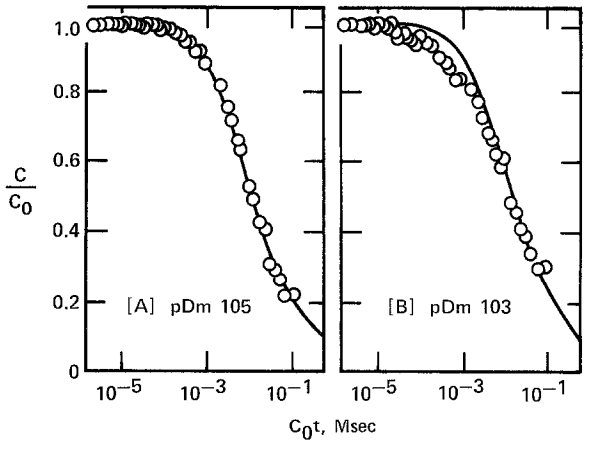

Figure 1. Self-Driven Reassociation Kinetics of pDm DNAs

The labeling of the pDm DNAs with ^{32}P and the reassociation reaction were carried out exactly as described by Wensink et al. (1974). C and C_o are the concentration of single-stranded DNA at time, t, and t = 0. Conditions for reassociation were: $C_o = 2 \times 10^{-7}$ M; 0.5 M NaCl in TE buffer; 65°C. The sigmoid curves are the least squares best fit of Equation (1) to the data (Morrow, 1974). The fitted curves for pDm101, 102, 105, 106, and 107 yield k_{pDm} values of 241, 258, 308, 339, and 163 M^{-1} sec^{-1}, respectively. The second order rate constants for the slow and fast reassociation components in pDm103 were determined as indicated in the text, and are 207 and 2×10^5 M^{-1} sec^{-1}, respectively.

in Figure 1A [k is the second order rate constant, and n = 0.44 when reassociation is assayed with S1 endonuclease (Morrow, 1974), as is the case here]. This is the expected result if no sequence repetition exists within the Dm segment, since in that case the pDm DNA should reassociate as a single kinetic component, given that the sequences in pSC101 are nonrepetitive (Wensink et al., 1974).

Such a fit is not, however, a very sensitive test for repetition, since the more rapid reassociation of Dm sequences caused by low frequencies of repetition will not be sufficient to produce detectable phases in the reassociation curve. A more sensitive test results from comparison of the observed value for the second order rate constant with that calculated from the equation

$$k_{pDm} = k_{pSC}(L_{pSC}/L_{pDm}) \quad (2)$$

which results from the fact the rate constants for nonrepetitious DNAs are inversely proportional to L, the length of their genomes (Morrow, 1974). The ratios of the observed to calculated k_{pDm}'s are 1.19 and 0.89 for pDm101 and 107, respectively, and are not significantly different from one. The ratios for pDm102, 105, and 106 are 1.44, 1.66, and 1.67, respectively. These higher observed k_{pDm}'s suggest that some sequence repetition may occur in the corresponding Dm segments. The frequency of that repetition cannot, however, be large. For example, one can calculate that if all sequences are repeated two times per Dm segment, the k_{pDm} will be increased 1.5–1.6 fold above that expected for no repetition.

The self driven reassociation kinetics for pDm103 is given in Figure 1B. The fall-off of the data points from the curve for Equation (1) that is seen at low values of Cot was reproduced in a second experiment, and indicates that a small fraction of the sequences in pDm103 reassociate faster than the rest. These sequences must be repeated within the Dm segment. Using the same procedure as that employed in the analysis of the biphasic curve shown in the inset to Figure 2A, we estimate:

—that approximately one tenth of pDm103 consists of sequences repeated on the order of 100 times per molecule;

—and that the remaining sequences are not repeated within the plasmid. That is, approximately 15% of the Dm103 segment, or approximately 3 kb, appears to consist of the repetition of a small sequence.

The repetition frequencies of Dm sequences within the D. melanogaster genome were determined from an analysis of the reassociation kinetics of a minute quantity of highly radioactive pDm DNA in the presence of a large amount of unlabeled D. melanogaster DNA, according to the method of Wen-

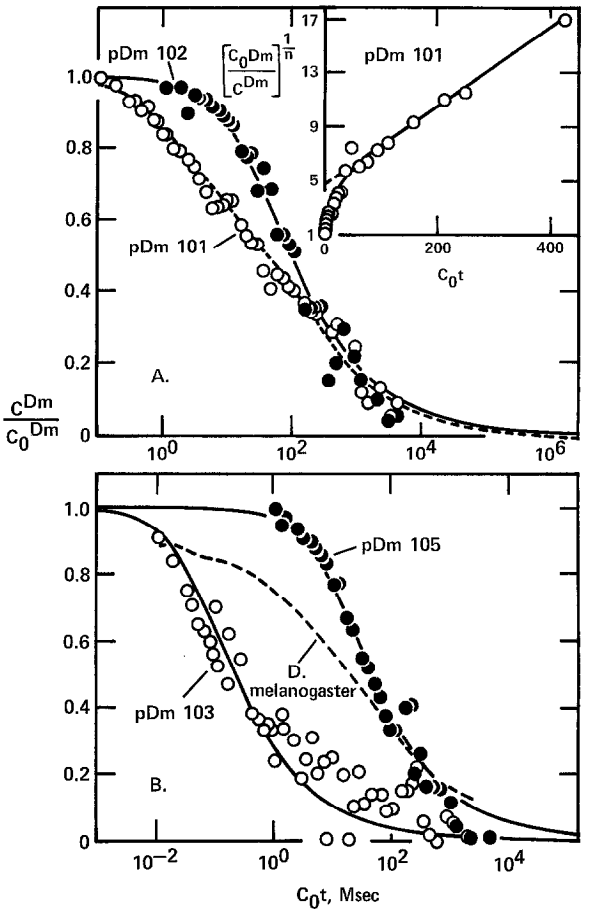

Figure 2. Reassociation Kinetics of Dm Sequences in the Presence of D. melanogaster DNA

A mixture of ^{32}P-labeled pDm DNA (approximately 2×10^{-9} M) and unlabeled D. melanogaster DNA (approximately 1×10^{-2} M) was sheared, denatured, and the reassociation of the labeled DNA followed as described by Wensink et al. (1974) under the conditions given in Figure 1. C^{Dm} and C_o^{Dm} are the concentrations of single-stranded DNA from the labeled Dm segments at t and t = 0. Values of C^{Dm}/C_o^{Dm} were obtained by subtracting out the contribution of the pSC101 sequences, as described previously (Wensink et al., 1974). The C_o in the abscissa refers to the unlabeled D. melanogaster DNA. The curves for Dm102, 103, and 105 are the least squares best fit of a modified form of Equation (1) to the data [that is, $C^{Dm}/C_o^{Dm} = (1 + k'_{pDm}C_ot)^{-n}$, with n = 0.44], and yield k'_{pDm} values of 0.039, 16.1, and 0.082 M^{-1} sec^{-1}, respectively; Dm106 and 107 (not shown) yield values of 0.100 and 0.036 M^{-1} sec^{-1}, respectively.

Rearrangement of this equation gives $[C_o^{Dm}/C^{Dm}]^{1/n} = 1 + k'_{pDm}C_ot$. A plot of $[C_o^{Dm}/C^{Dm}]^{1/n}$ compared against C_ot should and does yield a straight line for the above Dm segments; however, the biphasic curve shown in the inset is obtained for Dm101 and indicates it consists of two sequence classes with different repetition frequencies, α_1 and α_2. The portion of Dm101 occupied by each class, p_1 and p_2, and the α values were calculated from the following previously derived relations (Hogness et al., 1974): $p = 1 - I^{-n}$, $p_2 = I^{-n}$, $\alpha_1 = (k_1F_1/k_{pDm})(L_D/L_{pDm})$, $\alpha_2 = (k_2F_2/k_{pDm})(L_D/L_{pDm})$, where I is the intercept to the $[C_o^{Dm}/C^{Dm}]^{1/n}$ ordinate of the straight line formed by the data points at large C_ot values, k_1 and k_2 are the second order rate constants for the two classes, and F_1 and F_2 are the fractions of the D. melanogaster genome that they occupy. p_1 and p_2 were obtained directly from sink et al., 1974. The large amount of added D. melanogaster DNA controls and increases the rate of reaction of the labeled Dm sequences, but does not influence the reassociation rate of the labeled pSC101 sequences in the hybrid. The contribution of the pSC101 sequences to the labeled DNA that has reassociated can therefore be subtracted by computation to yield the reassociation kinetics of the Dm sequences that are shown in Figure 2.

The data points for Dm102, 105, 106, and 107 fit the second-order curve computed from Equation (1), as is exemplified by the data for Dm102 and 105 given in Figures 2A and 2B, respectively. If it is assumed that the Dm sequences are not repeated within the segment, then α, the repetition frequency of these sequences within the D. melanogaster haploid genome, is given the following equation (Wensink et al., 1974):

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

where k_{pDm} and L_{pDm} have their previously assigned meanings; k'_{pDm} is the rate constant for the reassociation of Dm sequences that is driven by total D. melanogaster DNA (Figure 2); and L_D is the length of DNA in the haploid genome, or 165,000 kb (Rudkin, 1972). The resulting values of α are given in group I of Table 2 and are all between 0.9 and 2.0.

If sequences are repeated within Dm102, 105, and 106, as suggested by the self driven kinetics, then their α values would be slightly larger. For example, if all sequences are equally repeated within the segment, then α is given by:

$$\alpha = (k'_{pDm}/k_{pSC})(L_D/L_{pSC}) \quad (4)$$

which yields the values given in parentheses in Table 2. Equation (4) is valid for any degree of repetition including none (Hogness et al., 1974). Equation (3) is preferred when no repetition exists within the segment, since both rate constants in the ratio, k'_{pDm}/k_{pDm}, are determined with the same preparation of labeled plasmid DNA.

The other two Dm segments are placed in another group in Table 2 because they contain sequences with repetition frequencies more than an order of

I and both equal 0.5. The values for k_1F_1 and k_2F_2, and hence α_1 and α_2, were determined by successive approximation. Starting with $k_2F_2 = S/I$, where S = slope of the line with intercept, I, k_1F_1 was calculated at low C_ot values from

$$C^{Dm}/C_o^{Dm} = p_1(C^1/C_o^1) + p_2(C^2/C_o^2) \quad (5)$$

where $C^1/C_o^1 = (1 + k_1F_1C_ot)^{-n}$ and $C^2/C_o^2 = (1 + k_2F_2C_ot)^{-n}$; see Hogness et al. (1974). The resulting value of k_1F_1 was then used to convert the observed C^{Dm}/C_o^{Dm} values to C^2/C_o^2, and a second approximation of k_2F_2 obtained from the slope of the line given by a plot of $(C_o^2/C^2)^{1/n}$ compared against C_ot. The resulting value (1.25×10^{-2} M^{-1} sec^{-1}) was similarly employed to obtain $k_1F_1 = 1.19$ M^{-1} sec^{-1}. Substitution of these and the above values for p_1 and p_2 into Equation (5) yields the dashed curve given in Figure 2A.

Table 2. Repetition Frequency and Chromosomal Location of the Dm Sequences

Dm Segment	α, Repetition Frequency[a]	Location[b]
Group I:		
Dm102	0.9 (1.3)	X: 8A
Dm107	1.3	2R:41D
Dm105	1.6 (2.7)	3R:82F
Dm106	2.0 (3.3)	2R:41D
Group II:		
Dm101		Chromocenter
Class 1, 50%	33	
Class 2, 50%	0.3	
Dm103	400	Nucleolus

[a] α was calculated from Equation (3), except for:
—the values given in parentheses, where Equation (4) was used,
—and those for the two sequence classes in Dm101, where the procedure given in the legend to Figure 2 was used.
[b] See Lindsley and Grell (1968), for numeration of the chromomeric regions, and Wensink et al. (1974), for the autoradiographic resolution.

magnitude greater than those in group I. Dm101 yields a biphasic curve (inset, Figure 2A), which is most simply interpreted as the result of two classes of sequences with different repetition frequencies. Analysis by a modification (Hogness et al., 1974) of a method described by Morrow (1974) and summarized in the legend to Figure 2, indicates that each class occupies about half of the Dm101 segment and that they exhibit α values of 0.3 and 33 repeats per genome (Table 2).

The quantitative analysis of Dm103 has proved difficult due to an exceptional scatter of the data points (Figure 2B) that is reproduced in repeated experiments. This scatter would prevent detection of the two classes of sequences in Dm103 that were indicated by the self driven kinetics (Figure 1B). We have therefore calculated an average value of α for the Dm103 sequences from Equation (3), where k'_{pDm} was computed from the least squares best fit of Equation (1) to the data points in Figure 2B. The value obtained is 400 repeats per haploid genome.

The Dm Segments Divide into Two Groups According to the Chromosomal Location of Their Sequences

The sequences in each Dm segment were mapped in the polytene chromosomes by in situ hybridization. The experiments were performed as described in the legend to Figure 3, using ^{3}H-labeled cRNA obtained by transcription of each pDm DNA in vitro with E. coli RNA polymerase, and ^{3}H-labeled DNA obtained by nick translation of pDm101, 102, 106 and 107 with E. coli DNA polymerase I. Examples of the labeling patterns are given in Figure 3, and the chromosomal sites obtained from such patterns are listed in Table 2. The results obtained with both cRNA and nick translated DNA were concordant. Earlier experiments (Wensink et al., 1974) demonstrated that the sequences in pSC101 DNA exhibit no in situ hybridization with D. melanogaster chromosomes.

All Dm segments consisting of sequences that exhibit little or no sequence repetition (group I, Table 2) are localized to a single chromomeric region. Although Dm106 and 107 are easily distinguished by their other properties and show no detectable homology in heteroduplex experiments, they exhibit the same pattern at the level of resolution of our in situ hybridization experiments, that is, to within a few bands. This was confirmed by comparing the pattern produced by a mixture of nick-translated DNAs from pDm106 and 107 (Figure 3d), to that produced by the individual DNAs. No difference in the pattern obtained from the three sets of slides was observed.

The sequences in Dm101 and 103, which contain moderately repetitive sequences (group II, Table 2), are not found in the chromomeric arms of the polytene chromosomes. Dm101 sequences are localized in the heterochromatin of the chromocenter, and Dm103 sequences in the nucleolus.

Dm103 Contains Sequences Complementary to 18S and 28S rRNAs

The nucleolar location of Dm103 sequences and their high repetition frequency suggest that Dm103 contains some part of the tandemly repeated gene-spacer units in the rDNA of Drosophila (Hamkalo, Miller, and Bakken, 1973; Meyer and Henig, 1974). Proof that this is the case has been obtained by hybridizing labeled 18S and 28S rRNA to nitrocellulose filters containing denatured pDm103 DNA (Figure 4). Under conditions of rRNA saturation, 7.9% of the pDm103 DNA on the filter hybridizes to the 28S rRNA, and 12.4% to a mixture of 18S and 28S rRNA. These results indicate that 4.2 kb of the pDm103 molecule corresponds to 28S rRNA sequences, and 2.3 kb correspond to 18S rRNA sequences. Since the lengths of these two rRNAs are 4.1 kb and 2.1 kb in Drosophila (Loening, 1968), we conclude that the Dm103 segment contains approximately one copy of each of the corresponding sequences.

Discussion

Group I Segments

The sequences in a group I segment are confined to a single chromomeric region and exhibit little or no repetition. These properties are inconsistent with tandem repetition models of the chromomeric

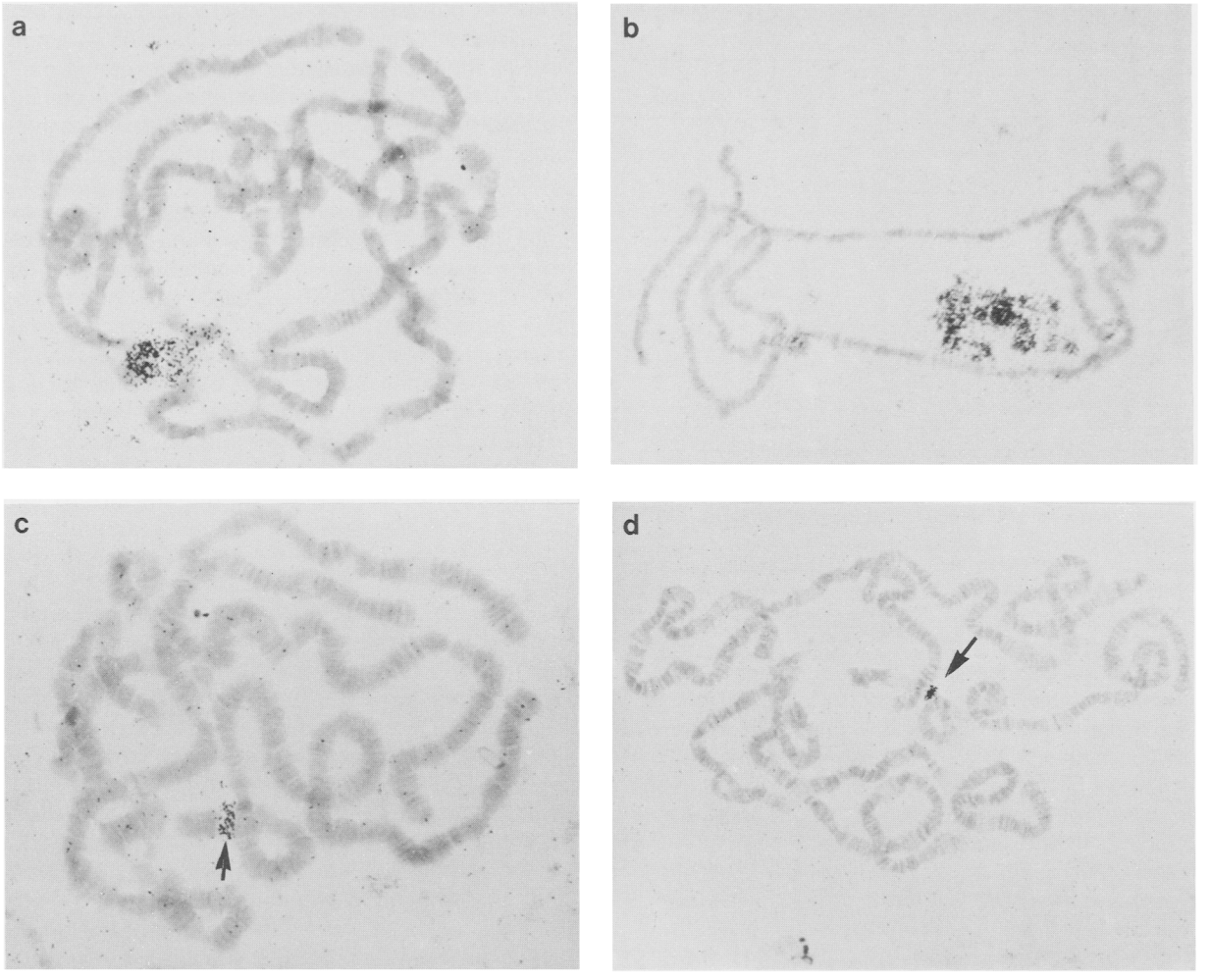

Figure 3. Hybridization of Dm Sequences to Polytene Chromosomes In Situ

Experiments were performed as described by Wensink et al. (1974), except that in some cases ^{3}H-labeled pDm DNA, prepared by nick translation (Schachat and Hogness, 1973), was used instead of cRNA.

(a) Labeling of α and β heterochromatin in the chromocenter by pDm101 ^{3}H–DNA (0.8 $\times$ 10^7 cpm/ml); 8 days exposure.

(b) Labeling of the nucleolus by pDm103 ^{3}H–cRNA (4.4 $\times$ 10^7 cpm/ml); 9 days exposure.

(c) Labeling of the 82F region in 3R (that is, right arm of chromosome 3) by pDm105 ^{3}H–cRNA (6.9 $\times$ 10^7 cpm/ml); 4 days exposure.

(d) Labeling of 41D in 2R by a mixture of ^{3}H–DNA from pDm106 (1.1 $\times$ 10^7 cpm/ml) and pDm107 (0.7 $\times$ 10^7 cpm/ml); 13 days exposure.

unit (Callan, 1967; Thomas et al., 1973). In these models, all or most of the unit is imagined to consist of a tandemly repeated sequence that codes for a single polypeptide. Such sequences are therefore expected to have an average length of 1–2 kb and be repeated 1–2 dozen times. The repetition frequencies given by the α values for group I segments are considerably smaller than this expectation (Table 2). In addition, the self driven reassociation kinetics indicate that either there is no substantial sequence repetition within the segment (Dm107), or, if present, is equivalent to only a 2-fold repetition of all sequences in the segment (Dm102, 105, 106). Hence, if these segments are part of or consist of tandemly repeated units, such units would be $\geq$ 18 kb in the case of Dm107, and $\geq$ approximately 8 kb for Dm102, 105, and 106.

Two of the three Dm segments examined earlier (Wensink et al., 1974) also fall into group I. The fact that group I segments appear with the same frequency in both sets suggests that this kind of segment represents most of the chromomeric DNA. This generalization is tentative because six of the nine segments analyzed to date were formed by EcoR1 cleavage, which, though partial, may introduce some selection in this sampling procedure. By

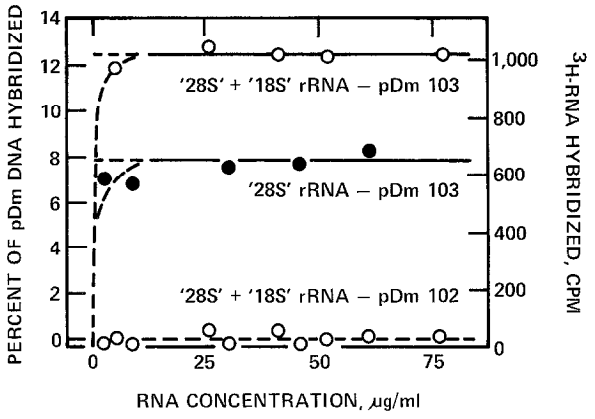

Figure 4. Hybridization of rRNA to pDm103 DNA

^{3}H ribosomal RNA (5.7×10^4 cpm/μg, see Experimental Procedures) was hybridized to ^{32}P-labeled pDm103 DNA (1.58×10^4 cpm/μg), or to the control pDm102 DNA (1.3×10^4 cpm/μg), on Schleicher and Schuell B6 filters essentially as described by Gillespie and Spiegelman (1965), except that the solvent was 50% formamide in 5 × SSC (0.4 ml) and the temperature was 55°C. The mixture of "28S" and "18S" rRNA was 1.6:1.0 by mass, respectively. Each data point is the mean of three determinations, normalized to 0.133 μg DNA/filter.

contrast, the first three segments were formed by shear breakage (Wensink et al., 1974), and are expected to result from random sampling.

The reassociation kinetics of group I segments obtained in the presence of D. melanogaster DNA give no indication of the short period interspersion of repetitive and nonrepetitive sequences observed in Xenopus laevis and sea urchins (Davidson et al., 1973). The mean lengths of the repetitive and nonrepetitive sequence elements in this short period interspersion are 0.3 kb and 0.7–1.1 kb, respectively. The repetitive sequences therefore account for 20–30% of the pattern, and this amount should be detected by the kinetics illustrated in Figure 2, provided that the repetitive sequences are comparably defined with $\alpha > 20$. We should not expect to detect an interspersion pattern in which such repetitive sequences represent $< 10\%$ of the segment, or in which they are defined to include sequences repeated only a few times per genome.

Our results are, in fact, consistent with the pattern of interspersion observed in D. melanogaster (Manning et al., 1975), in which moderately repetitive sequences of number average length 5.6 kb are separated from each other by repetitive sequences of length greater than 13 kb.

Group II Segments

Both Dm101 and 103 contain sequences with repetition frequencies an order of magnitude or more greater than those for sequences in group I segments, and these moderately repetitive sequences are located in the heterochromatin. Dm101 is about equally divided between sequences with $\alpha = 33$ and those with $\alpha = 0.3$. The finding of an α value near 0.25 suggests that Dm101 is a segment of the Y chromosome. This is because the D. melanogaster DNA used to drive the reassociation reaction derives from a 1:1 mixture of males and females, and therefore a nonrepetitive sequence in Y, X, or an autosome will exhibit an α value of 0.25, 0.75, or 1.00, respectively. Furthermore, the Y chromosome is confined to the heterochromatic chromocenter of the polytene chromosomes, where the Dm101 sequences map (Figure 3). Whether the moderately repetitive sequences and nonrepetitive sequences are interspersed within this segment or occur in single blocks can be determined by analysis of its EcoR1 fragments (Table 1), or of fragments produced by other restriction endonucleases. This kind of internal sequence mapping has been reserved for a subsequent phase of these studies.

The location, repetition frequency, and rRNA complementarity of the sequences in Dm103 all indicate that it derives from the nucleolus organizer (NO) in either the X or the Y chromosomes (Ritossa et al., 1966). The NO in the X chromosome of D. melanogaster (Oregon R) contains about 260 repeats of the 18–28S rRNA gene, and that in the Y chromosome somewhat less (Spear, 1974). The α value of 400 that we obtain for Dm103 is not significantly different from these repetition frequencies, given the scatter of the data points (Figure 2).

These genes occur in tandem arrays of gene-spacer units that have been visualized in the electron microscope as transcription matrices (Hamkalo et al., 1973; Meyer and Henig, 1974; Chooi and Laird, personal communication). The simplest assumption is that Dm103 was cut from the middle of such an array by EcoR1, and, since it contains no internal EcoR1 cleavage sites (Table 2), is equal to or less than a repeating unit. Given that the length of the 38S primary transcript is 8.3 kb (Perry et al., 1970), then the length of the spacer region would be ≥ 8.8 kb. The length of the "spacer chromatin" in the transcription matrices observed by Chooi and Laird is about 1.4 μm and 0.8 μm for the X and Y chromosomes, respectively. Given a foreshortening of up to 50% for the DNA in such spacer regions (Laird, Foe, and Chooi, personal communication), these lengths represent up to about 8.4 kb and 4.8 kb of DNA in the X and Y spacers, respectively. This suggests that Dm103 represents the repeating unit in the NO of the X chromosome, and that the length of the repeating unit in the NO of the Y chromosome is about 13 kb (that is, 4.8 + 8.3 = 13.1 kb). This suggestion is

strengthened by the recent observation that agarose gel electrophoresis of D. melanogaster DNA that had been completely digested with EcoR1 yielded two major length classes of fragments capable of hybridizing to cRNA from pDm103—one of about 17 kb and other of approximately 12 kb (unpublished experiment of D. M. Glover).

Experimental Procedures

Bacteria

E. coli K12 strain HB101 is *hsm*$^-$, *hrs*$^-$, *recA*$^-$, *gal*$^-$, *pro*$^-$, *str*^{R} (Boyer and Roulland-Dussoix, 1969) and was obtained from H. W. Boyer; C600 (pSC101) was obtained from S. N. Cohen (Cohen et al., 1973).

DNAs and Enzymes

D. melanogaster (Oregon R) DNA, and the pSC101 and pDm plasmid DNAs were isolated as described by Wensink et al. (1974). EcoR1 endonuclease was prepared according to the procedure of Green et al. (1974). The E. coli enzymes, DNA ligase (Modrich, Anraku, and Lehman, 1973), DNA polymerase I (Jovin, Englund, and Bertsch, 1969), and RNA polymerase (Berg, Barrett, and Chamberlin, 1971) were prepared according to the indicated references and generously provided by P. Modrich, D. Brutlag, and W. Wickner, respectively. DNAase I and Aspergillus oryzae S1 endonuclease were obtained from Worthington Biochemical Corp. and Seikagaku Kogyo Co., respectively.

Electron Microscopy and Buoyant Densities

DNA lengths were measured in the electron microscope relative to reference DNAs (pSC101 = 9.2 kb, Wensink et al., 1974; λ = 46.5 kb, Davidson and Szybalski, 1971) present on the same grid prepared by aqueous spreading (Davis, Simon, and Davidson, 1971). Buoyant densities of DNAs in CsCl were determined as described by Champoux and Hogness (1972).

Construction of Hybrid Molecules

The pDm plasmids were constructed by the following modification of the EcoR1-ligase method (Mertz and Davis, 1972; Cohen et al., 1973). D. melanogaster DNA was partially digested with EcoR1 endonuclease and then fractionated according to length by electrophoresis (2 volts/cm, 36 hr, 20–25°C) in 0.7% agarose slab gels (Green et al., 1974). The fragments in individual slices were recovered by electrophoresis and their lengths determined by electron microscopy. 40% of the fragments in the selected fraction (14.2 ± 2.3 kb; see text) have EcoR1 termini at both ends as judged by their ability to cyclize in the presence of E. coli DNA ligase (Mertz and Davis, 1972). This DNA was mixed with EcoR1-cleaved pSC101 DNA (Wensink et al., 1974) so that the ratio of fragments with two EcoR1 termini to pSC101 molecules was 1.3, and the joining reaction was carried out in two stages. The DNA (130 μg/ml) was incubated in 0.1 mM DPN, 1 mM EDTA, 10 mM each of $(NH_4)_2SO_4$, $MgSO_4$, and Tris–HCl (pH 7.5), 14 μg/ml DNA ligase, 100 μg/ml BSA for 45 or 110 min at 14°C. It was then diluted 33 fold in the above reaction buffer, more ligase added to a final concentration of 13 μg/ml, and incubation at 14°C continued for 16 hr. The products were fractionated by zone sedimentation in a sucrose gradient (Hobom and Hogness, 1974) to exclude circular hybrids larger than $(Dm)_1(pSC)_1$. Because the standard deviations of the lengths of circular DNAs in the relevant gradient fractions was 4–5 kb, elimination of $(Dm)_2(pSC)_1$ hybrids (expected lengths = 37.6 ± 3.3 kb) required the exclusion of fractions in which the mean length was greater than 20–22 kb. Purification and yield of the desired $(Dm)_1(pSC)_1$ hybrids (expected lengths = 23.4 ± 2.3 kb) were therefore sacrificed in favor of eliminating the undesired hybrids when combining fractions. The DNA in the combined fractions was used for the transformation steps described in the text.

Preparation of ^{3}H Ribosomal RNA

^{3}H-labeled rRNA (5.7 × 10^4 cpm/μg) was prepared from Schneider's line 2 of D. melanogaster cells grown in spinner cultures as described previously (Blumenthal, Kriegstein, and Hogness, 1973), except that the media contained 10 μCi/ml ^{3}H-uridine (29 Ci/mmole) at inoculation (4 × 10^5 cells/ml), and 20 μCi/ml at 48 hr. Cells were harvested at a density of 2 × 10^6 cells/ml, washed twice in isotonic saline (9 mg/ml NaCl, 0.42 mg/ml KCl, 0.25 mg/ml $CaCl_2$, 0.2 mg/ml $NaHCO_3$, 2.5 mg/ml glucose), suspended in the aqueous phase obtained from an equilibrium mixture (0°C) of phenol containing 0.1% 8-hydroxyquinoline and 150 mM NaCl, 1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 7.4), 10 μg/ml polyvinylsulphate, and disrupted in a Dounce homogenizer. After addition of the phenol phase, shaking and separation of phases by centrifugation, the aqueous phase was made 0.1% in sodium dodecylsulphate and 0.2% in diethylpyrocarbonate, shaken again with phenol, and the resulting aqueous phase layered onto a 5–20% sucrose gradient (0.2 M NaCl in TE). After centrifugation (Beckman SW41 rotor, 5 hr, 40,000 rpm, 4°C), the peak fractions were combined to yield, after dialysis and ethanol precipitation, the "18S" rRNA, whereas the combined fractions at the leading edge of the "28S" peak were ethanol precipitated, centrifuged as above, and the peak fractions from this second gradient combined, dialyzed, and concentrated to yield the "28S" rRNA.

Acknowledgment

This work was supported by research grants from the National Science Foundation and the National Institutes of Health. D. M. G. was a Damon Runyun fellow; R. L. W. and D. J. F. are Jane Coffin Child fellows.

Received February 28, 1975

References

Beermann, W. (1962). 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*, 81–86.

Blumenthal, A. B., Kriegstein, H. J., and Hogness, D. S. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 205–223.

Boyer, H. W., and Roulland-Dussoix, D. (1969). J. Mol. Biol. *41*, 459–472.

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

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

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

Davidson, E. H., Graham, D. E., Neufeld, B. R., Chamberlin, M. E., Amenson, C. S., Hough, B. R., and Britten, R. J. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 295–301.

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

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.

Gillespie, D., and Spiegelman, S. (1965). J. Mol. Biol. *12*, 829–842.

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.

Hamkalo, B. A., Miller, O. L., and Bakken, A. H. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 915–919.

Hobom, G., and Hogness, D. S. (1974). J. Mol. Biol. *88*, 65–87.

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

Hogness, D. S., Wensink, P. C., Glover, D. M., White, R. L., Finnegan, D. J., and Donelson, J. E. (1974). In The Eukaryotic Chromosome, W. J. Peacock and R. D. Brock, eds. (Canberra: Australian National University Press), in press.

Jovin, T. M., Englund, P. T., and Bertsch, L. (1969). J. Biol. Chem. *244,* 2996–3008.

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

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

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

Loening, U. E. (1968). J. Mol. Biol. *38,* 355–365.

Manning, J. E., Schmid, C. W., and Davidson, N. (1975). Cell *4,* 141.

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

Meyer, G. F., and Henig, W. (1974). Chromosoma *46,* 121–144.

Modrich, P., Anraku, Y., and Lehman, I. R. (1973). J. Biol. Chem. *248,* 7495–7501.

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

Nathens, D., and Smith, H. O. (1975). Ann. Rev. Biochem. *44,* 273–293.

Perry, R. P., Cheng, T.-Y., Freed, J. J., Greenberg, J. R., Kelley, D. E., and Tartoff, K. D. (1970). Proc. Nat. Acad. Sci. USA *65,* 609–616.

Rittossa, F. M., Atwood, K. C., Lindsley, D. L., and Spiegelman, S. (1966). Nat. Cancer Inst. Monogr. *23,* 449–472.

Rudkin, G. T. (1972). In Results and Problems in Cell Differentiation, *4,* W. Beermann, 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.

Spear, B. B. (1974). Chromosoma *48,* 159–179.

Thomas, C. A., Pyeritz, R. E., Wilson, D. A., Dancis, B. M., Lee, C. S., Bick, M. D., Huang, H. L., and Zimm, B. H. (1973). Cold Spring Harbor Symp. Quant. Biol. *38,* 353–370.

Thomas, M., and Davis, R. W. (1975). J. Mol. Biol., in press.

Wensink, P. C., Finnegan, D. J., Donelson, J. E., and Hogness, D. S. (1974). Cell *3,* 315–325.