Cell, Vol. 10, 177-192, February 1977, Copyright © 1977 by MIT

R Loop Mapping of the 18S and 28S Sequences in the Long and Short Repeating Units of Drosophila melanogaster rDNA

Raymond L. White* and David S. Hogness†
Department of Biochemistry
Stanford University School of Medicine
Stanford, California 94305

Summary

Cleavage of D. melanogaster rDNA with the Eco R1 restriction endonuclease reveals two major classes of repeating units: a long class of 17 kilobases (kb) and a short class of 11.4 kb. R loop mapping has been used to determine the topography of the sequences corresponding to the 18S and 28S rRNAs in both classes, including a cloned member (Dm103) of the long class. This mapping procedure derives from a novel reaction that we discovered between single-stranded RNA and homologous regions in duplex DNA molecules. In high formamide and at an elevated temperature, the RNA pairs with one of the two DNA strands in the region of homology to form an R loop in which one element is an RNA/DNA duplex and the other is single-stranded DNA (Figure 1). Mapping is accomplished by visualization of R loops in the electron microscope. The R loop map of Dm103 parallels that determined independently by Glover and Hogness (1977) from an analysis of its restriction fragments. Both maps indicate that the 28S rDNA in this cloned unit is divided into two blocks by a 5 kb insertion segment. R loop mapping of a population of long units obtained directly from the rDNA (that is, without cloning) has demonstrated that this interruption of the sequence coding for the 28S rRNA is a characteristic of the long class. By contrast, the 28S rDNA in the short units that we examined is not interrupted by an insertion segment. Otherwise, the R loop maps of the long and short units do not differ significantly. The two classes therefore correspond to repeating units that do (IN^+) or do not (IN^-) contain an insertion segment. Models for the transcription and function of these two classes of repeating untis are discussed.

Introduction

When duplex DNA is mixed with single-stranded RNA in the presence of formamide and at an elevated temperature, R loops will form if the nucleotide sequence in the RNA is homologous to that in an internal segment of the DNA. Figure 1 shows a diagram of this reaction and defines the structure of an R loop. If the homologous region in the DNA is terminal, a fork will be formed in which one prong is an RNA/DNA duplex and the other is a single DNA strand, like the two elements in the loop.

We discovered this reaction during a search for a method that would allow the fractionation of duplex DNA molecules that contain sequences homologous to a given RNA from those that do not, while maintaining the association between the two strands in the desired DNA. Our purpose was to facilitate the cloning of specific eucaryotic genes in bacteria by decreasing the complexity of the population of DNA molecules to be inserted into the cloning vector. The model system that we used in approaching this problem consisted of the pDm103 hybrid plasmid described in the accompanying paper (Glover and Hogness, 1977), and of the D. melanogaster 18S and 28S rRNAs that hybridize to the 17kb repeating unit of rDNA contained in this hybrid. In our first attempts to anneal these RNAs to the rDNA, we used supercoiled plasmids, on the theory that the negative twists would stabilize the R loop, much as is the case for D loops (Wang, 1974). We carried out the annealing reaction in the presence of 50% formamide because Birnstiel, Sells, and Purdom (1972) had shown that the melting temperature of an RNA/DNA duplex is higher than that of the corresponding DNA duplex in such a solvent. In fact, R loops do form in such supercoils. To our surprise, however, we found that if the formamide concentration was raised to 70%, while maintaining the temperature at 40°C, R loops are also formed in the open circular plasmid and in the linear Eco R1 fragment that represents the long rDNA unit (that is, Dm103). Since then, Thomas, White, and Davis (1976) have shown that the maximum rate of R loop formation in linear DNA occurs at the denaturation temperature of the DNA segment forming the R loop, indicating that the reaction is initiated by pairing between the RNA and a single strand of DNA in a partially denatured molecule.

Although we and other investigators (Wellauer and Dawid, 1977; R. W. Davis, personal communication) have used the R loop reaction for DNA fractionation, the fact that R loops are easily visible in the electron microscope led us to develop first a method whereby they can be used to map the region in a DNA duplex that is homologous to a given RNA. In this paper, we have applied that method to map the 18S and 28S sequences, not only in Dm103, but also in the uncloned Eco R1 fragments that represent the two major classes of rDNA units in D. melanogaster, which exhibit mean lengths of 16.6 and 11.4 kb. The R loop map of Dm103 parallels, at higher resolution, that obtained by restric-

---

* Present address: Department of Microbiology, University of Massachusetts, School of Medicine, Worcester, Massachusetts 01605.

† To whom requests for reprints should be addressed.

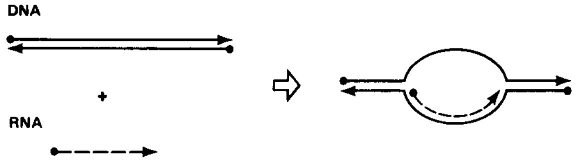


Figure 1. Diagram of the R loop Reaction

The black dot and arrowhead represent the 5′ and 3′ termini, respectively.

tion mapping (Glover and Hogness, 1977) and provides an independent demonstration that the 28S rDNA in this unit is interrupted by a 5 kb insertion segment. This unusual arrangement is also seen in the R loop map of the uncloned long units, indicating that it is a general characteristic of this class. By contrast, the map of the short units demonstrates that the sequences coding for the 28S rRNA in this class are contiguous. Thus the long and short units correspond to those that do (IN^+) or do not (IN^-) contain an insertion segment in the 28S rDNA, and it is this difference that we attempt to explain in the models for the function of these two classes that are presented at the end of this paper.

Results

The Expected Topographies of R Loops in Dm103

When 18S and/or 28S rRNA are mixed with Eco R1-digested pDm103 in 70% formamide buffer [70% formamide, 0.1 M NaCl, 0.01 M EDTA, 0.1 M Tris-HCl (pH 7.4)] and this mixture is then incubated at 40°C, the partially denatured Dm103 molecules react with these rRNAs to form R loops. We first consider the topographies of the R loops that are predicted from the previously constructed map of the 18S and 28S rDNAs in Dm103 (Glover and Hogness, 1977), and then show that the observed topographies conform to these predictions.

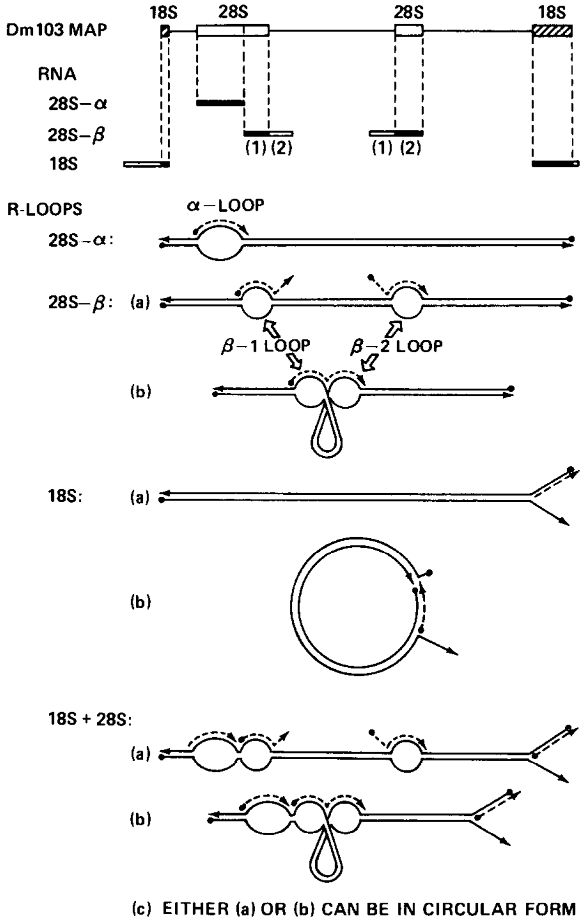


Figure 2. Expected R Loops Formed by Single Molecules of Dm103

The map of Dm103 was constructed by Glover and Hogness (1977) from an analysis of the amount of 18S and 28S rDNAs in ordered restriction fragments of Dm103. The expected sequence homologies between Dm103 and rRNAs are indicated by the filled portion of the blocks that represent the indicated rRNA molecules, each of which is assumed to be 2 kb long. The 28-α and 28-β symbols represent the α and β fragments released upon denaturation of the 28S rRNA (see text). The sequence in the β fragment is divided into two parts (β-1 and β-2) which are indicated by (1) and (2) in the diagram. Two identical β fragments are given to show that the β-1 and β2 sequences are separated in Dm103. Similarly, two identical 18S rRNA molecules are shown to indicate the sequence homologies at each end of Dm103. The symbols used in the R loop diagrams are the same as those used in Figure 1. The orientation of the RNAs is that expected from the direction of transcription of D. melanogaster rDNA (see Discussion).

The map is shown at the top of Figure 2. The 28S rDNA in Dm103 is divided into two blocks of approximately 3 kb and 1 kb that are separated by about 5 kb of DNA of unknown function. Our predictions about the R loops formed by the 28S rRNAs are based on the assumption that the arrangement of the 28S rDNA in Dm103 is equivalent to that which would be obtained if this 5kb DNA segment were inserted at the appropriate position in a continuous 4 kb stretch of 28S rDNA whose nucleotide sequence matched that in the 4kb 28S rRNA molecule. Such a continuous stretch of 28S rDNA is found in the shorter 11.4 kb repeating units described at the end of this paper, and we presume that transcription of these stretches yields the 28S rRNA used in our studies (see Discussion). This 28S rRNA was isolated from D. melanogaster cell cultures, where it is first formed as a continuous 4 kb polynucleotide chain (Jordan, Jourdan, and Jacq, 1976). However, a central hidden break soon appears in this chain so that the vast majority of the isolated 28S rRNA consists of two fragments which are so loosely coupled that they dissociate under denaturing conditions milder than those given above for the formation of R loops (Greenberg, 1969; Shine and Dalgarno, 1973; Jordan et al., 1976). The central location of the break was deduced from the observation that the length of each fragment is about equal to that of the 2 kb 18S rRNA. Given the above assumption regarding the arrangement of the 28S rDNA in Dm103, the sequence corresponding to one of the fragments, called α, must be located at the left end of the large block of 28S rDNA, whereas the sequence corresponding to the other fragment, called β, must be divided into two parts: β-1, located adjacent to the α sequence in the large block, and β-2, corresponding to sequence in the small block. These sequence homologies are indicated in Figure 2.

Figure 2 also shows the expected sequence homologies between the single-chain 18S rRNA and the blocks of 18S rDNA located at each end of Dm103. This expectation results from the conclusion that the Dm103 repeating unit was cut out from the chromosomal rDNA by cleavage at the Eco R1 site in the 18S rDNA block of each of two adjacent repeating units (Glover and Hogness, 1977).

The simplest of the R loop topographies expected from these sequence homologies is that for the single R loop formed by the α fragment (Figure 2). By contrast, two R loops can be formed by the β fragment, and these can appear in more than one configuration. When a single β fragment forms either a β-1 or a β-2 R loop, the unpaired portion of the fragment should form an RNA "whisker" that projects from the right end of the β-1 loop or from the left end of the β-2 loop. If, on the other hand, a single β fragment pairs with both the β-1 and β-2 regions to create two R loops in the same Dm103 molecule, then these whiskers should be absent and the 5 kb insertion segment should form a third loop consisting solely of duplex DNA. We call this configuration a "triple loop," and it is shown in Figure 2 immediately below the whiskered β-1 and β-2 R loops. The observation of these R loop configurations would clearly demonstrate that the insertion segment interrupts a contiguous region that codes for the β fragment. In addition, we expect to find structures in which a single β fragment forms a β-1 R loop in one Dm103 molecule and a β-2 R loop in another, thereby forming a bridge between the two DNAs (not shown).

Pairing between the 18S rRNA and the 18S rDNA at the right end of Dm103 will create a fork that should have a small RNA tail at the end of one prong. Pairing of this tail with the 18S rDNA at the left end of the molecule will yield a circular structure with short and long DNA whiskers separated by an RNA/DNA duplex with a length equal to the 18S rRNA. Alternatively, the small RNA whisker may pair with another Dm103 molecule to yield linear DNA dimers (not shown).

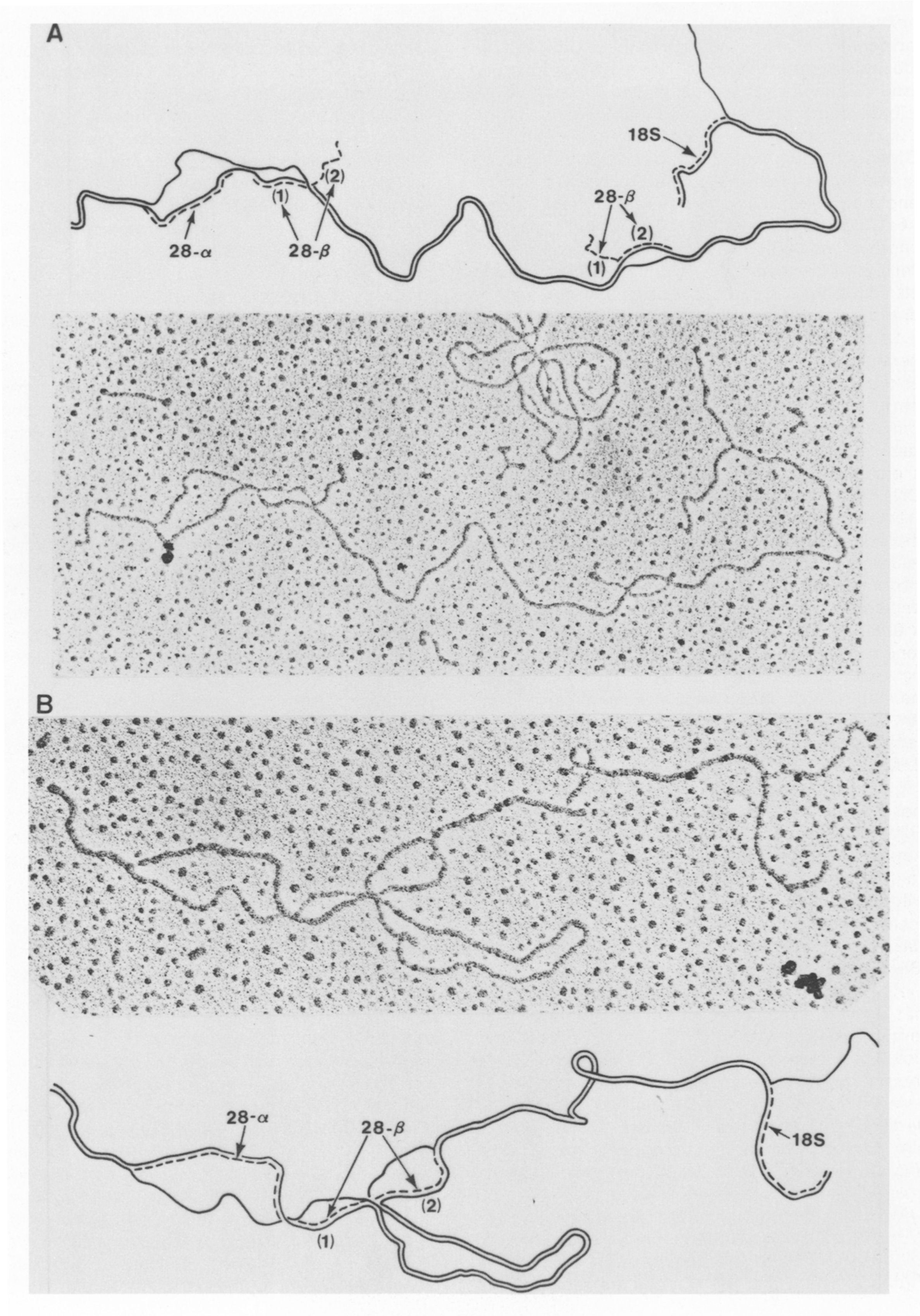
Finally, the topographies that are expected when a Dm103 molecule is saturated with both 18S and 28S rRNA R loops are shown at the bottom of Figure 2. The α and β-1 R loops are separated in the diagram, although they may combine to form a larger R loop. As will become apparent, the configuration of these two R loops will depend not only upon the nature of the break in the 28S rRNA that yields the α and β fragments (that is, whether it involves a single cleavage event or multiple cleavages that result in the loss of nucleotides at the breakpoint), but also upon the methods used to visualize the R loops in the electron microscope. It should be noted that denaturation of the 28S rRNA also yields two very small fragments—the 5.8S and the 2S rRNAs (Jordan et al., 1976). These are not treated here because their combined length is only 0.18 kb, and they have not been detected in R loops.

The Observed Topographies of R Loops in Dm103

Each of the R loop topographies predicted above has been observed in the electron microscope. When the annealing mixture contains a molar excess of all rRNAs relative to the Dm103 DNA, one observes Dm103 molecules that are saturated with R loops. Two representatives of this saturated class are shown in Figure 3, where it can be seen that their configurations conform to those given at the bottom of Figure 2.

The fork at the right end of these molecules is produced by the 18S rRNA. This has been demonstrated by annealing purified 18S rRNA to Dm103. The predominant product of this reaction is the forked molecule shown in Figure 4A. These molecules exhibit a mean length of 17.1 ± 0.18 kb (N = 22), which equals the length of Dm103 (Glover et al., 1975; in this paper, the standard error for N measurements is given with the mean). Given that the single-stranded prong of the fork is DNA, then its mean length is 1.64 ± 0.028 kb (N = 20), which should equal the length of the 18S rDNA segment at the right end of Dm103. The duplex prong should represent an 18S rRNA/DNA duplex with a very short RNA tail at the tip of the prong (Figure 2). Since such a tail was not convincingly observed, the entire prong was treated as if it were a duplex to obtain a mean length of 1.70 ± 0.044 kb (N = 21).

Although these forked monomers do not provide explicit evidence for such an RNA tail, its existence is demonstrated by two other structures that are frequently formed in the same reaction mixture. One of these is the circular monomer shown in Figure 4B (mean length = 16.6 ± 0.11 kb; N = 15), and the other is a linear dimer of Dm103 (mean length = 33.7 ± 0.32 kb; N = 4) that often has a fork at one end (not shown). Both structures exhibit the short and long whiskers that can be seen in Figure 4B. In the dimers, the short whisker is always closest to the forked end and is separated from it by a length equivalent to one Dm103 molecule. The du-


A
18S
28-β
(2)
(1)
28-α
28-β
(2)
(1)
B
28-α
28-β
(2)
(1)
18S

plex segment between the whiskers must therefore represent an RNA/DNA joint connecting the opposite ends of Dm103, and its length (1.84 ± 0.033 kb) should equal that of the 18S rRNA.

If the length of the single-stranded prongs in the forked molecules is taken as the length of the 18S rDNA segment at the right end of Dm103, then the difference between the above whisker-to-whisker length and this value [that is, 1.84 (± 0.033) kb − 1.64 (± 0.028) kb = 0.20 (± 0.043) kb] equals the length of the 18S rDNA segment at the left end of Dm103. Alternatively, the lengths of the right and left 18S rDNA segments can be obtained from the lengths of the long (1.72 ± 0.032 kb) and short (0.12 ± 0.021 kb) whiskers (N = 15 circular monomers), respectively. The fractional standard error for the short whiskers is large because the uncertainty in determining the position of the unattached terminus of the whisker represents a substantial fraction of its length. The fractional error, however, is equally large when the length of the left 18S rDNA segment is determined from the difference between two much greater lengths. Since there is also little reason to weight differently the two values for the right 18S rDNA segment, we have taken the simple average of the values for the lengths of the right and left segments (that is, 1.68 ± 0.021 kb and 0.16 ± 0.025 kb, respectively) in constructing the R loop map of Dm103 shown in Figure 5A. These two values are to be compared to the lengths of the right (1.62 kb) and left (0.27 kb) 18S rDNA segments obtained in the accompanying paper, where the fractional standard errors are about 9% (Glover and Hogness, 1977).

The three remaining R loops exhibited by the molecules in Figure 3 are formed by the two 28S rRNA fragments, and are called the α loop, the β-1 loop, and the β-2 loop, reading from left to right. Each of these loops was observed after annealing only the 28S rRNA to Dm103 under conditions identical to those given in the legend to Figure 4, except that the 18S rRNA was replaced with 3.0 μg/ml of 28S rRNA. The resulting molar concentration of the α and β fragments (2.3×10^{-9} M) is about one third that for Dm103 (6.3×10^{-9} M), and under these conditions, the probability of observing Dm103 molecules annealed to only one fragment is increased over that when the RNA is in excess. Many molecules contain only the α loop, whose duplex element exhibits a mean length of 1.75 ± 0.026 kb (N = 11), and which is separated from the left end of Dm103 by a duplex DNA segment (called L) with a length of 1.44 ± 0.019 kb (N = 20) (Figure 5A). Subtracting the 0.16 ± 0.025 kb of 18S rDNA at the left terminus of Dm103 from L, we find that the length of the DNA separating the 18S and 28S rDNA is 1.28 ± 0.031 kb.

The duplex elements of the β-1 and β-2 loops exhibit mean lengths of 0.74 ± 0.011 kb (N = 13) and 1.21 ± 0.014 kb (N = 25). The fact that the sum of these lengths (1.95 ± 0.018 kb) is approximately one half the 28S rRNA and is significantly different from the length of the α loop is one of the reasons for concluding that both loops are formed from the other, or β fragment of 28S rRNA. Another reason is the fact that when the loops appear individually or in the unlinked state shown in Figure 3A, they are always accompanied by a whisker at one end—the right end of the β-1 loop and the left end of the β-2 loop. By contrast, when the β-1 and β-2 loops are linked, either in the triple loop structure shown in Figure 3B, or in side-to-side dimers of Dm103 in which the β-1 loop in one Dm103 is linked to the β-2 loop in the other (not shown), then neither loop exhibits a whisker. Hence we conclude that the whisker attached to an unlinked β-1 loop contains the sequences that are paired in the β-2 loop, and reciprocally, that the whisker at the β-2 loop contains the 28S rRNA sequences that are paired in the β-1 loop. The lengths of these RNA whiskers should therefore correspond to the lengths of the two loops. Although we have not studied the lengths of the whiskers in detail because of the high variance that they exhibit, measurement of the contour lengths of five whiskers of each type indicates that the whisker at the β-2 loop is about half the length of the whisker at the β-1 loop (0.5 ± 0.06), where the expected ratio is 0.61 ± 0.012. (If we assume that these RNA whiskers exhibit the same length/kb ratio as single-stranded DNA, then they contain 0.6 ± 0.05 kb and 1.2 ± 0.09 kb, in approximate agreement with the expected values of 0.74 ± 0.011 kb and 1.21 ± 0.014 kb). Given that the β-1 and β-2 loops are both formed by the β fragment, then the position of the whiskers attached to each loop indicates that the respective 28S rDNA segments in Dm103 must have the same orientation.

The mean length of the duplex DNA that separates the β-1 and β-2 loops is 5.32 ± 0.068 (N = 16), and is taken from measurements on both linear and triple-loop molecules, since the values obtained for each class are not significantly different. The duplex DNA between the right end of the β-2 loop and

Figure 3. Dm103 Saturated with 18S and 28S R Loops

18S rRNA (3.8 μg/ml; 5.8×10^{-9} M) and 28S rRNA (2.7 μg/ml; 2.0×10^{-9} M) were mixed with Eco R1-digested pDm103 (22 μg/ml; 1.2×10^{-9} M in Dm103) in 70% formamide buffer and incubated at 40°C for 4.0 hr. The molar concentrations for the 18S and 28S rRNAs were calculated from the molecular lengths obtained in this work. The length of pDm103 used to calculate the molar concentration of DNA is 26.3 kb (Glover et al., 1975). Samples were prepared for electron microscopy by method A (Experimental Procedures).

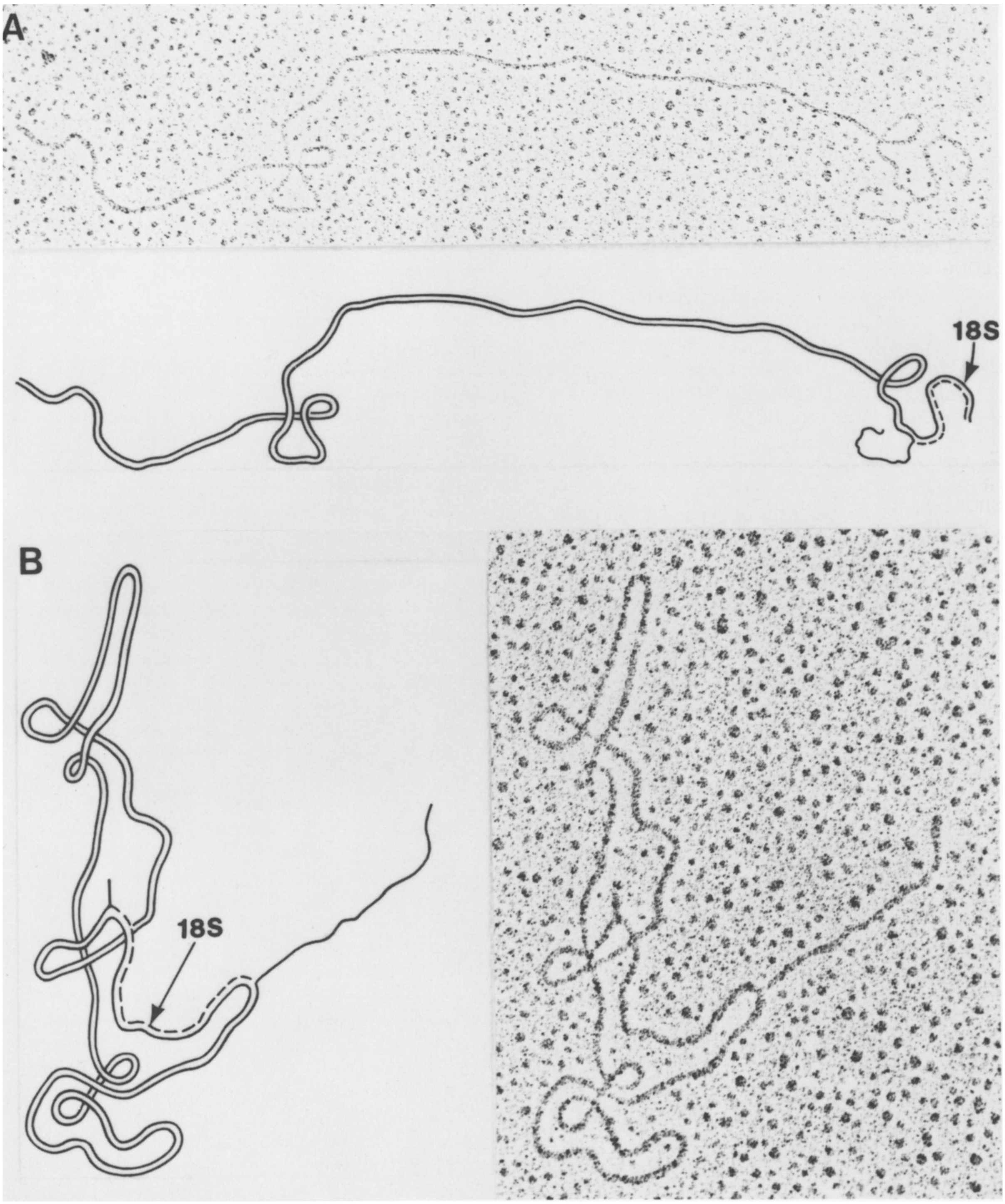


Figure 4. 18S R Loops in Dm103

18S rRNA (4.2 µg/ml; 6.4 × 10^{-9} M) was mixed with Eco R1-digested pDm103 DNA (110 µg/ml; 6.3 × 10^{-9} M in Dm103) in 70% formamide buffer and incubated at 40°C for 4 hr. Samples were prepared for microscopy by method B (Experimental Procedures).

the right terminus of Dm103 in molecules that do not contain the 18S rRNA fork is 6.10 ± 0.024 kb (N = 20). Subtraction of the 1.68 ± 0.021 kb of 18S rDNA at the right terminus of Dm103 from this value yields a length of 4.42 ± 0.032 kb for the DNA separating the 28S and 18S rDNAs (called R) at this end of the molecule (Figure 5A).

Finally, we have attempted to determine the

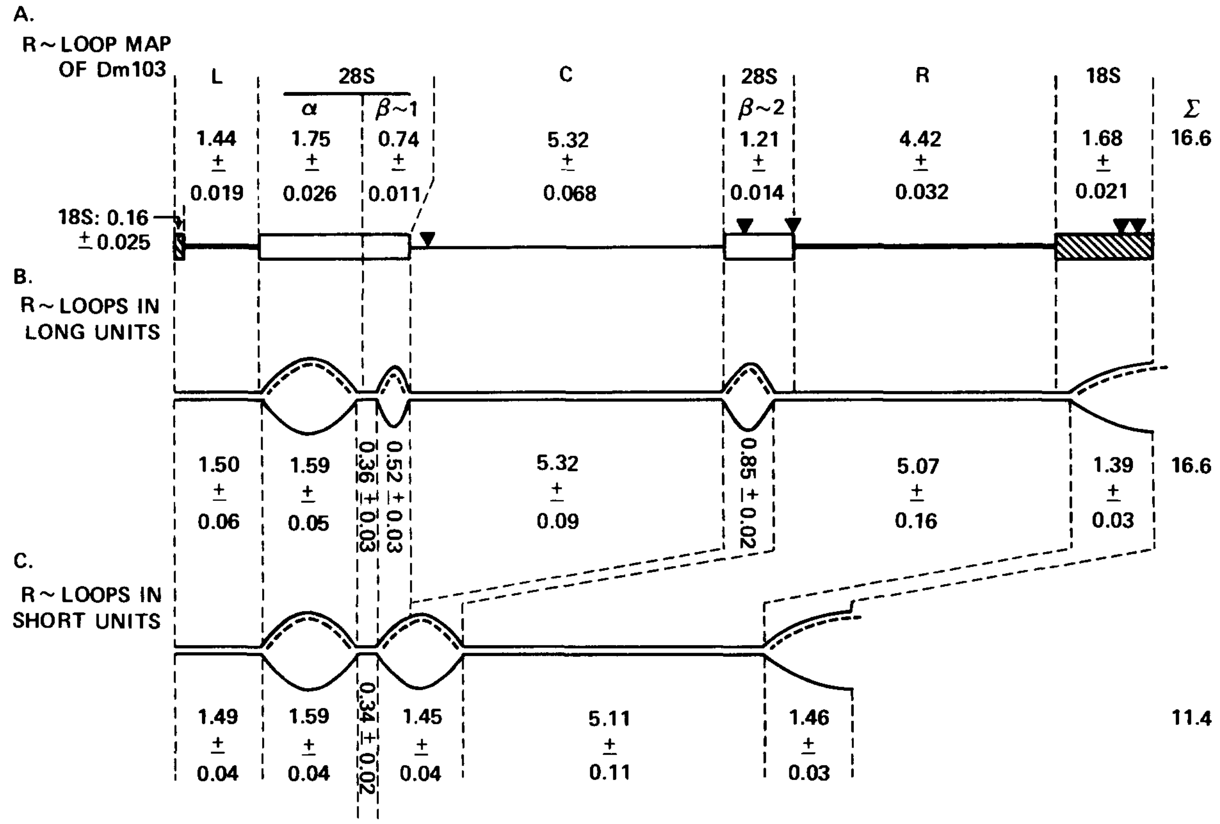


Figure 5. R Loop Topographies of Dm103 and the Uncloned Units

(A) The mean lengths and standard errors given in kb units for each map interval (see text for N) were taken from electron micrographs prepared by method B from mixtures of Eco R1-cut pDm103 (110 μg/ml; 6.3 × 10^{-9} M) and 18S rRNA (4.2 μg/ml; 6.4 × 10^{-9} M) or 28S rRNA (3.0 μg/ml; 2.3 × 10^{-9} M) in 70% formamide buffer which had been annealed at 40℃ for 6 hr. The solid triangles indicate the positions of the Hind III sites when the restriction map (Glover and Hogness, 1977) and this R loop map are normalized to the same total length.

(B) and (C) Fractions enriched for the indicated uncloned units (Experimental Procedures) were annealed with 18S rRNA or 28S rRNA and prepared for electron microscopy as described in Figure 7. The lengths (± standard errors) for the intervals were obtained in the same manner as described in the text for Dm103, and N = 14 ± 2 for each interval. Standard errors (SE; standard deviation of the mean) given here and in the text were calculated as described by Davis, Simon, and Davidson (1971).

amount of DNA, if any, that separates the α and β-1 28S rDNA sequences at the left end of Dm103. Although the α and β-1 loops are clearly separated when the molecules are prepared for electron microscopy as indicated in Figure 3 (that is, by method A), this is not the case when they are prepared by the alternative method B that was used for all the length measurements given above (see Experimental Procedures for both methods). When method B is used, the R loops whose left ends are located at a distance L from the left terminus of Dm103 divide into two distinct classes. The first class consists of the α loops. The second class consists of longer loops (mean length = 2.55 ± 0.080 kb; N = 8) that have β-1-like whiskers at their right ends. Both the length and the attached whisker indicate that loops of the second class are formed by the combined annealing of the α fragment and the β-1 segment of the β fragment. The sum of the lengths of the α and β-1 loops is 2.49 ± 0.028 kb, and any DNA that separates the α and β-1 sequences should equal the difference between the length of the combined loop and this sum, or 0.06 ± 0.085 kb. Since this value is not significantly different from zero, no separation is indicated on the map in Figure 5A, although any separation ≤0.2 kb is compatible with the data.

To explain the observation that molecules prepared by method A, but not method B, exhibit a constriction between the α and β-1 loops, we note that when equivalent R loops obtained by both methods are examined, those obtained by method A are 10–30% shorter (dependent upon the class of loop) and sometimes have very short whiskers at their ends. (These whiskers are not to be confused with the much longer β-1 and β-2 whiskers, which are seen with both methods.) Apparently this shortening results from a displacement of the RNA at the ends of the RNA/DNA duplex by the free DNA strand through branch migration (Lee, Davis, and Davidson, 1970). We suggest that this same displacement mechanism acts to create the duplex

DNA constriction between the α and β-1 loops. We do not know what elements of method A promote this R loop shortening, but the lower concentration of formamide in the hypophase (A, 10%; B, 20%) and/or the slightly higher temperature of the secondary incubation between annealing and spreading (A, 28°C; B, 25°C) are obvious candidates.

Frequencies of the 28S R Loops in Dm103

We have examined the same annealed mixture of Dm103 DNA and 28S rRNA that was used for the above length measurements to determine the frequencies of the different kinds of Dm103 molecules that are formed under this condition of Dm103 excess. The results are given in Table 1, where the scored structures have been divided into six classes (column 1) whose frequencies are given in columns 2 and 3. The fraction of all Dm103 molecules in the annealed mixture that are present in each class (column 4) was then calculated from the class frequencies under the assumption that these adequately represent all the products of the annealing reaction (see Table 1, footnote a). The last two columns in the table give the fraction of all α and β fragments in the annealing mixture that are bound to the Dm103 molecules in each class.

Since the sums of the bound α and β fraction equal 0.99 and 1.03, respectively, we conclude that essentially all the 28S rRNA fragments have formed R loops after 6 hr of incubation at 40°C. This result confirms our previous conclusion that both 28S rRNA fragments form R loops in Dm103. This is satisfying, since our previous conclusion was based solely on the difference in lengths between the α loop and the sum of the β-1 and β-2 loops.

If the α loop is formed by one fragment and the

Table 1. Molecular Frequencies in an Annealed Mixture of Dm103 DNA and 28S rRNA

Class[a]		Class Frequencies		Dm 103 Molecules (Fraction of Total in Mixture)[b]	α and β RNA Molecules (Fraction of Total in Mixture)[c]	
		Number	%		α	β
(I)	Monomers without R Loops	172	42.3	0.393		
(II)	Monomers with:					
	Single R Loops					
	α[d]	89	21.9	0.203	0.56	
	β-1 or β-2[e]	40	9.8	0.091		0.25
	Multiple R Loops					
	$\alpha + \beta$-2[f]	20	4.9	0.046	0.13	0.13
	Triple Loop[g]	55	13.5	0.126	0.14	0.34
(III)	Dimers[h]	31	7.6	0.142	0.16	0.31
	$\Sigma =$	407	100.	1.00	0.99	1.03

[a] The conditions for annealing and electron microscopy are indicated in Figure 5A. All structures in random fields were examined. After elimination of the easily recognizable linear pSC101 molecules, 490 structures remained, of which 407 (83%) could be unambiguously scored as members of the six classes given in this column. Although scoring of the remaining 17% was sufficiently ambiguous that they are not included here, we have no reason to suppose that they include significant additional classes.

[b] This fraction was calculated by dividing the number of Dm103 molecules in each class (that is, the number of monomers or two times the number of dimers per class) by 438, the total number of Dm103 in the 407 scored structures.

[c] This fraction was computed by first determining the molar concentration of Dm103 in each class from the product of the Dm103 fraction (preceding column) times 6.3×10^{-9} M, the initial concentration of Dm103 in the mixture. The corresponding molar concentration of the α and β fragments were then obtained from the product of the Dm103 concentration times the average number of α and β fragments per Dm103 molecule in the structures within each class (see footnotes d–h). The indicated fractions were obtained by dividing the α and β class concentrations by 2.3×10^{-9} M, the initial concentration of α and of β in the annealing mixture.

[d] α/Dm103 = 1.0; β/Dm103 = 0.0. The fraction of molecules in this class that may contain an $\alpha + (\beta$-1) combined loop, instead of a simple α loop, is sufficiently small to be neglected without significantly affecting the total β frequency (see footnote g).

[e] α/Dm103 = 0.0; β/Dm103 = 1.0. Most (approximately 90%) are β-2 loops.

[f] α/Dm103 = 1.0; β/Dm103 = 1.0.

[g] α/Dm103 = 0.4; β/Dm103 = 1.0. Triple loops are formed when a single β fragment forms β-1 and β-2 loops in a single Dm103 (see text). The fraction of triple loops that contains an $\alpha + (\beta$-1) combined loop in place of the single β-1 loop was estimated to be 0.4 from a subset of 11 molecules whose loop lengths were measured. The number of these combined loops is much greater in this and the dimer class (footnote h) than in the preceding classes (where they can be neglected) because essentially all approximately (96%) of the β-1-containing loops are confined to the triple-loop and dimer classes, whereas the distribution of α-containing loops is not biased with respect to these classes.

[h] α/Dm103 = 0.4; β/Dm103 = 0.8. The (β-1)-(β-2) linkage that characterizes the Dm103 dimers is effected by a single β fragment, yielding a β/Dm103 ratio of 0.5. In addition, an average of about 0.3 β fragments per Dm103 are found in unlinked β-1 and β-2 loops. The α/Dm103 ratio of 0.4 results from $\alpha + (\beta$-1) combined loops and separate α loops. The frequencies for these additional α and β fragments were obtained from a subset of 8 measured dimers.

two β loops by the other, and if the sequences in the two fragments do not overlap, then we expect that the two classes of loops with be independently distributed among the Dm103 molecules. That this is the case is demonstrated by the following comparison between the expected and observed frequencies of the different kinds of Dm103 molecules in the annealed mixture. Given that all the α and β fragments are bound to Dm103, then the expected probability that a Dm103 molecule will contain an α fragment (p_α) equals the initial α concentration (2.3×10^{-9} M) divided by the initial Dm103 concentration (6.3×10^{-9} M), or 0.37. The observed fraction of Dm103 molecules that contain α can be obtained from Table 1 by reference to the appropriate footnotes and is 0.36 [that is, $0.203 + 0.046 + (0.4)(0.126) + (0.4)(0.142) = 0.36$]. The probability that a Dm103 molecule will contain a β fragment (p_β) is expected to be somewhat different than p_α, since a single β can be shared by two Dm103 molecules in the dimer configuration, and a Dm103 molecule within a dimer may contain an additional β (Table 1, footnote h). Using the dimer frequencies to correct the β and Dm103 concentrations for the affects of these two forms, we obtain $p_\beta = 0.42$. The observed fraction of Dm103 molecules containing β is 0.41 (that is, $0.091 + 0.046 + 0.126 + 0.142 = 0.41$).

Using 0.37 and 0.42 for p_α and p_β, respectively, the expected fractions of Dm103 molecules that do not contain R loops [$(1\text{-}p_\alpha)(1\text{-}p_\beta)$], that contain only α loops ($p_\alpha - p_\alpha p_\beta$), that contain only β loops ($p_\beta - p_\alpha p_\beta$), and that contain both α and β loops ($p_\alpha p_\beta$) are 0.37, 0.21, 0.26, and 0.16. The respective observed values are 0.39, 0.20, 0.25, and 0.15. It should be emphasized that the effect of correcting p_β for dimers is small; that is, if the uncorrected value for p_β of 0.37 is used, the expected fractions are 0.40, 0.23, 0.23, and 0.14, respectively.

In contrast to the independence exhibited by the α and β loop classes, the distribution of β-1 and β-2 loops is consistent with their formation from a common fragment. For example, the β-2 loop appears much more frequently than the β-1 loop in monomers containing only one of these loops (Table 1, footnote e). This is the expected result if both loops are formed by different segments of the β fragment whose lengths correspond to the lengths of the respective loops. Other things being equal, we expect that the probability of forming β-2 loops from free β fragments will be greater than for β-1 loops because of the greater length of the β-2 segment (Figure 5), and that the half-life of monomers containing the β-2 loops will be greater than that for monomers with only β-1 loops because the probability that the shorter RNA whisker attached to the β-2 loop will form a β-1 loop, either in the same molecule (triple loop monomers) or in a different molecule (dimers), should be less than that for the reciprocal reactions resulting from the longer whisker attached to the β-1 loop. The observation that essentially all the Dm103 molecules that contain both a β-1 and a β-2 loop are confined to the triple loop and dimer classes is also consistent with the conclusion that both loops are formed from the same RNA fragment.

The Topographies of R Loops in Repeating Units That Have Not Been Cloned

The only Eco R1 cleavage site in the repeat unit represented by Dm103 is within the block of 18S rDNA. If this characteristic applies to most or all of the repeating units within the rDNA, then Eco R1 digestion of total D. melanogaster DNA should yield a set of rDNA fragments whose members represent the individual repeating units. In particular, if these units are distributed in a small number of major length classes, each class should form a band after electrophoresis of the Eco R1 digest in agarose gels. Figure 6 shows that there are only two major length classes of rDNA units, an observation that we noted in a previous publication

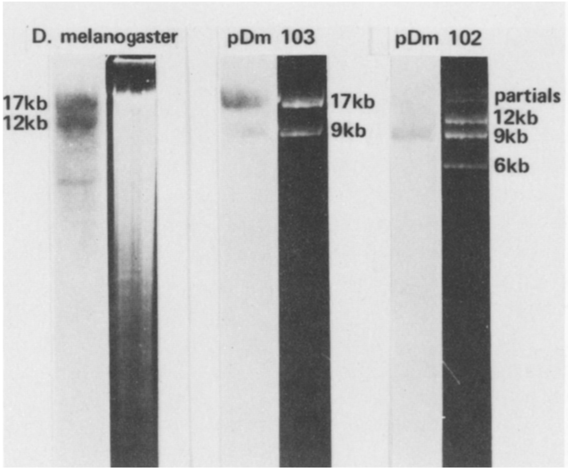


Figure 6. Southern Hybridization of ^{32}P-Labeled cRNA from pDm103 to Eco R1 Fragments of D. melanogaster DNA

Total embryonic DNA, as well as purified pDm103 and pDm102 plasmid DNAs, were digested to completion with Eco R1 endonuclease and fractionated by electrophoresis in 0.7% agarose gels as described by Glover and Hogness (1977). The fluorescence of the DNA stained with ethidium bromide is seen in the right-hand panel of each pair, and the autoradiogram of its hybridization to the ^{32}P-cRNA from pDm103 is shown in the left-hand panel. Hybridization was carried out as described by Southern (1975). As expected, both the Dm103 (17 kb) and pSC101 (9 kb) fragments of pDm103 hybridize with this cRNA. Eco R1 digestion of pDm102 yields linear pSC101 plus two fragments from the Dm102 segment (6 kb and 12 kb; Glover et al., 1975), neither of which contains rDNA. Hybridization is therefore expected and seen only with the pSC101 DNA. The lengths of the two major bands of hybridized fragments from the embryonic D. melanogaster DNA were obtained by comparison to these control panels.

(Glover et al., 1975). The fragments of rDNA in the digest of embryonic D. melanogaster DNA were detected by hybridization of the DNA in the gel to ^{32}P-labeled cRNA transcribed in vitro from pDm103 DNA. The hybridization was carried out according to Southern's method (1975), and the resulting autoradiograph reveals major bands of rDNA at 17 kb and 11-12 kb. Autoradiographs that have been exposed for longer times contain several minor bands, one of which is visible at about 5 kb in Figure 6. We are concerned here, however, only with the long and short fragments in the two major bands.

We have examined the R loops formed between the purified rRNAs and these uncloned fragments to answer two questions: first, does the class of large fragments contain molecules that exhibit the same topography as Dm103, and second, do the small fragments exhibit the more typical topography in which the 28S rDNA sequences are contiguous? A positive answer to the first question would quiet our concern that the peculiar arrangement of the 28S rDNA in Dm103 might have arisen from a rearrangement of its DNA during cloning; and a positive answer to the second question would identify the probable source of the 28S rRNA.

Since R loop mapping requires that only one of the two reactants be purified, the Eco R1 fragments were only enriched for the long and short rDNA units by zone sedimentation prior to formation of R loops with purified 18S and 28S rRNAs (Experimental Procedures). The molecules were prepared for electron microscopy by method A so that we could more easily detect the adjacent α and β loops when on the same molecule (see above). Molecules in the fraction enriched for the long fragments exhibited all the R loop configurations detected for Dm103 (Figures 2-4). Because of the limited nature of the question posed above, we selected for contour measurements those molecules that exhibited a similar topography to that found for Dm103. It should be emphasized, however, that these represent a majority of the molecules that contain R loops in this fraction, and that this qualitative visual selection was made before, not after, the contour lengths were measured.

The results of these length measurements are shown in Figure 5B. Except for the shorter lengths of the R loops that were expected from the use of method A, the observed topography clearly parallels that for Dm103 (see Discussion). Although not indicated in this figure, these molecules also contain a small amount of 18S rDNA at the left end, as is evidenced by the formation of circular molecules of the type shown in Figures 2 and 4B. We therefore conclude that the peculiar topography of Dm103 does not arise from a rearrangement of its DNA, and that the majority of rDNA units in this fraction exhibit its topography.

The majority of molecules that form R loops in the fraction enriched for the smaller rDNA fragments exhibit a quite different topography that is shown in Figure 7. Such molecules were selected for contour length measurement as indicated above, with results that are shown in Figure 5C. The significant difference between the short and long fragments of rDNA is that the insertion segment, C, is missing from these short fragments so that a single loop is formed by the β fragment of the 28S rRNA. This conclusion is confirmed by the observation that the length of this single β loop is, within experimental error, equal to the sum of the lengths of the β-1 and β-2 loops for the long units. The R loop maps of the long and short fragments are otherwise virtually identical. In particular, the separation between α and β-1 loops in the long fragments is equivalent to that observed between the α and β loops in the short fragments, when both were prepared for microscopy by method A. We therefore presume that the amount of DNA, if any, that separates the α and β sequences in the short fragments is the same as that determined for the separation of the α and β-1 sequences in Dm103—that is, ≤ 0.2 kb. If such a small separation does indeed exist, it can be assumed to result from the loss of a small segment of the primary 28S rRNA when it is cleaved to yield the α and β fragments (Jordan et al., 1976).

Finally, we note that the 18S rRNA not only creates forks at the right end of the short fragments (Figures 5C and 7A), but also forms circular molecules that are identical to those observed with Dm103 (Figure 4B), except that the circular contour lengths are shorter by about 5 kb. This observation demonstrates that the left end of the short fragments also contains a small segment of 18S rDNA, from which we infer that the short fragments, like the long, represent complete repeating units of rDNA.

Discussion

The R Loop Map of Dm103

The R loop map of the 18S and 28S rDNA in Dm103 given in Figure 5A confirms and extends the resolution of that obtained from restriction fragments (Glover and Hogness, 1977). Both maps indicate that Dm103 is a complete repeating unit of rDNA, cut out by cleavage at an Eco R1 site in the 18S rDNA block. Within this unit, the 28S rDNA is divided into two blocks that are separated by an insertion segment, C, that is 5.3 kb in length. The larger of these two blocks consists of all the sequences in the 1.75 kb α fragment and 38% of the

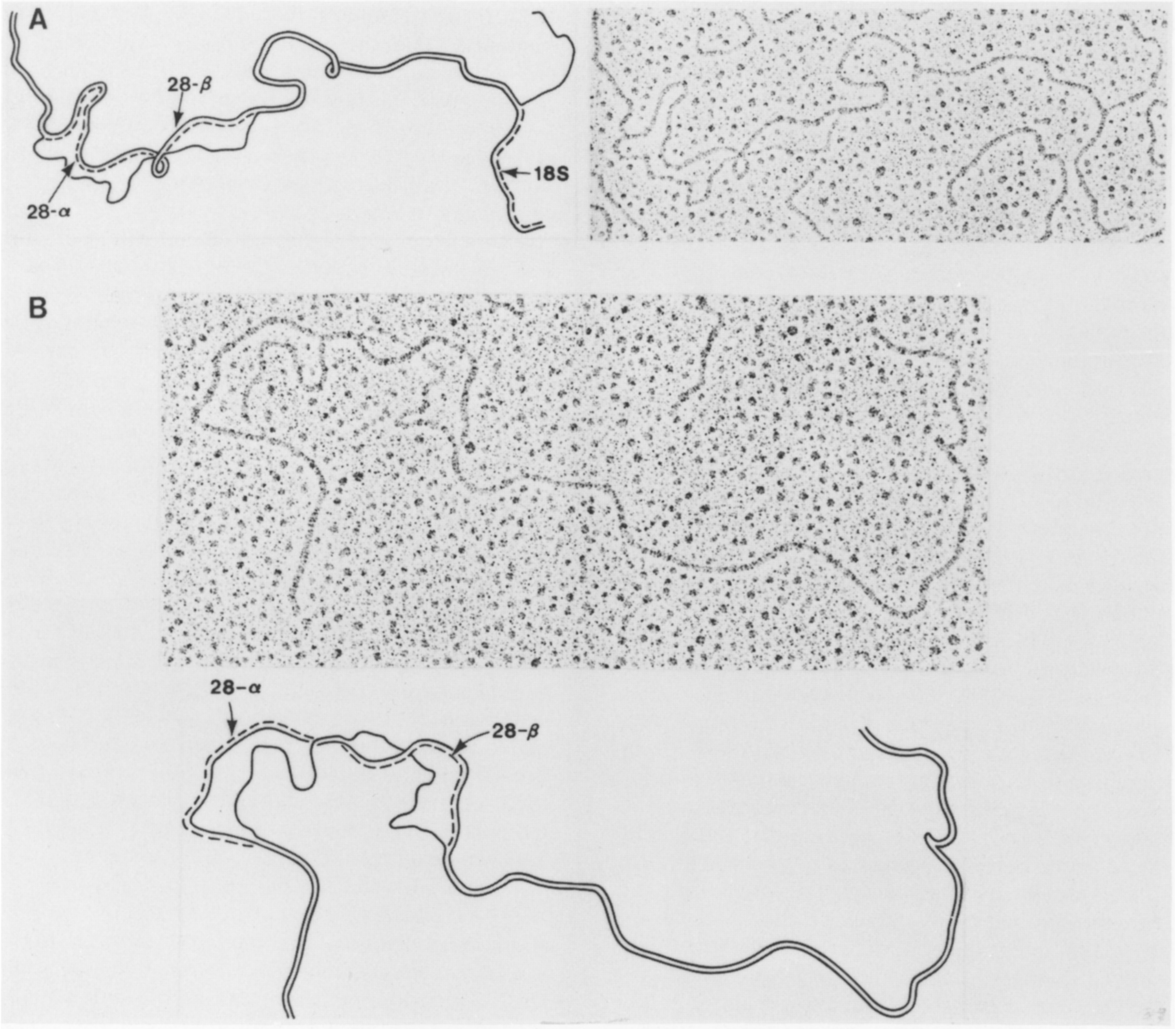


Figure 7. R Loops Formed by the Short Uncloned Units

DNA from fractions of Eco R1 fragments of embryonic DNA that were enriched for short (or long) units (Experimental Procedures) was mixed with an excess of 28S and/or 18S rRNA in 70% formamide buffer and annealed for 12 hr at 40°C. Samples were prepared for microscopy by method A, except that open circular pSC101 DNA and RNAase (0.1 μg/ml) were added with the cytochrome c immediately before spreading to provide a length reference and to reduce the background of unreacted RNA in the electron micrographs, respectively. The molecules shown above derive from the fraction enriched for short units and have formed R loops with both 18S and 28S rRNAs (A) or only with 28S rRNA (B). Molecules enriched for the long units form 18S and 28S R loops (not shown) like those formed in Dm103 (Figures 3 and 4), as described in the text.

sequences in the 1.95 kb β fragment. The remaining 62% of the sequences in the β fragment occupy the shorter block of the 28S rDNA.

How accurately do these R loop measurements reflect the lengths of the RNA molecules that produce them? Previous length determinations of Drosophila 18S and 28S rRNAs were obtained by electrophoresis in nondenaturing gels (Loening, 1968; Perry et al., 1970; Meyer and Hennig, 1974) and by filter hybridization of these rRNAs to Dm103 DNA (Glover et al., 1975; Glover and Hogness, 1977). A value of 4.1 kb was obtained from each of the three electrophoretic measurements of the 28S rRNA—in good agreement with the mean and range of 4.15 (± 0.05) kb obtained from the two hybridization experiments. These values include the 0.18 kb in the small 2S and 5.8S fragments (Jordan et al., 1976), and this must be subtracted before they are compared to the 3.70 ± 0.032 kb obtained for the sum of the α and β fragments by R loop mapping of Dm103 (Figure 5A). This comparison indicates that the R loops are 6.2 (±0.6)% shorter than would be expected from the electrophoretic and hybridization values. The length of the 18S rRNA determined by R loop mapping (1.84 ± 0.033 kb; Figure 5A) is also about 6% shorter than that expected from the mean of the electrophoretic values, which range from 1.9-2.0 kb. It is also shorter than the mean value of the hybridization measurements for the 18S sequences, but their larger range (1.9-2.3 kb) makes this comparison less meaningful.

We do not know whether the shorter lengths of the R loops result from defects in the technique, or from defects in the electrophoretic and hybridization measurements. However, some uncertainties in the R loop measurements should be noted. In calculating the lengths of R loops in kb units, it was assumed that the number of base pairs per μg of contour length is the same for RNA/DNA and DNA/DNA duplexes. Although the available evidence indicates that this assumption is valid to within ±7% (Experimental Procedures), the remaining uncertainty is sufficient to explain the shorter lengths.

The length of the R loop might also be shorter than the length of the corresponding RNA molecule if the ends of the RNA, rather than pairing with the DNA, formed minute tails projecting from the termini of the loop. The fact that such tails are not visible in the electron micrographs used for the above length measurements (that is, those prepared by method B; Experimental Procedures) is not a sufficient argument against this possibility. Should these tails account for 6% of the RNA in, for example, the α loop, it is improbable that they would be detected, since their mean length would be only 0.05 kb. The existence of even minute tails would account for our observation that the RNA in R loops can be completely digested with pancreatic RNAase under conditions where the R loops are otherwise maintained. One need only postulate that the tails are attacked by the RNAase and that they are continuously generated throughout the digestion by branch migration (Lee et al., 1970). We have obtained evidence supporting this postulate by demonstrating that prior treatment of the R loops with the single-strand-specific endonuclease, S1, makes the RNA/DNA duplex element of the loop RNAase resistant, presumably by destroying the single-stranded DNA and thereby preventing branch migration. Indeed, the RNAase resistance provided by the S1 treatment has been used as a convenient assay for R loop formation (R. L. White, unpublished experiments).

We have also demonstrated that observable tails can be formed at the ends of R loops by changing the conditions of sample preparation (for example, as in method A), and that this results in a consequent shortening of the R loop lengths. This shortening effect is clearly seen in Figure 5, where the sum of the lengths of the 28S rRNA R loops is decreased from 3.70 ± 0.032 kb when method B is used (Figure 5A), to 2.96 ± 0.06 kb (Figure 5B) or 3.04 ± 0.06 kb (Figure 5C) when method A is used—a shortening of 19%. The length of the 18S rRNA fork is similarly decreased by 15% with the same change in methods. While none of these observations demonstrates that the R loop lengths obtained with method B are indeed shorter than the corresponding RNAs, they do indicate that defects in the R loop technique will probably result in shorter rather than longer lengths.

The topography of R loops in the Uncloned Units

A large fraction, at least a majority of the Eco R1 rDNA fragments in the fraction that was enriched for the long units, exhibits the R loop topography given in Figure 5B. Comparison with the R loop map of Dm103, after accounting for the effects of method A on this topography, indicates that the structure of the uncloned units is very similar if not identical to that for Dm103. Thus the increase observed in the length of the R segment of the uncloned units (5.07 − 4.42 = 0.65 kb; Figure 5) is accounted for by the combined shortening of the β-2 loop and the 18S fork [(1.21-0.85) + (1.68-1.39) = 0.65 kb], provided that shortening of β-2 is restricted from proceeding at the end adjacent to the insertion segment, C. If the shortening of the β-1 loop is similarly restricted, then the agreement between the observed lengths of the insertion segments is also indicative of equivalence between Dm103 and the uncloned units. Such a restriction is consistant with the fact that the distance from the left end of Dm103 to the (β-1)/C boundary (3.99 ± 0.08 kb) equals the corresponding distance in the uncloned units (3.97 ± 0.09 kb).

The length distribution observed for the insertion segments in the uncloned long units (SD = 0.32 kb) is essentially the same as that for the homogeneous Dm103 population (SD = 0.27 kb). Since these uncloned units were subject to visual selection for the majority type within a fraction that was enriched for the long units, we obviously cannot conclude that all insertion segments exhibit the length of that in Dm103. We are convinced, however, that inserts of this length constitute the most frequent class in our source of rDNA, since few if any of the units in the other major fraction of rDNA contained insertion segments. Thus the majority of short units in this fraction exhibit the topography given in Figure 5C. This topography can be directly compared with that for the uncloned long units and, except for the absence of an insertion segment, is seen to include all the regions defined for the long units without significant change in their lengths.

We therefore conclude that the major variable in the embryonic rDNA units is the presence or absence of an insertion segment at a specific locus in the 28S rDNA. Subsequent measurements by Wellauer and Dawid (1977) indicate a length heterogeneity among these insertion segments. Their observations on the rDNA in the 17 kb band are consistent with ours in indicating that most insertion segments in these long units belong to one length class of about 5 kb, although they also detected minor length classes ranging between 4.5 and 6 kb in this fraction. The frequency of these minor classes is sufficiently low that they might not have been included in the smaller selected population that we examined. They also observed shorter insertion segments among the minor classes of rDNA units that are found between the two major bands (Figure 6), as well as some very short insertions of about 0.5 kb in a minority of the units within the 11-12 kb band—insertions that we again may not have observed because of our selective procedures. Pellegrini, Manning, and Davidson (1977) have also observed a heterogeneity in the lengths of the insertion segments that is somewhat different and is peculiar in that it does not conform to that expected from the distribution of the long units seen by Eco R1 cleavage and gel electrophoresis (Figure 6; Tartof and Dawid, 1976; Wellauer and Dawid, 1977). Perhaps this difference results from a difference in source of the rDNA or some selection during the isolation of the rDNA. For example, Pellegrini et al. (1977) isolated their rDNA from organisms after puparium formation, whereas we and Wellauer and Dawid (1977) obtained the rDNA from embryos. Similarly, Wellauer and Dawid used a stock that contained a bobbed mutation in the Y chromosome, and Tartoff and Dawid (1976) have found that bobbed mutations in the X chromosome cause an additional heterogeneity in the rDNA, as observed by electrophoresis of Eco R1 fragments.

It should be emphasized that the existence of such a heterogeneity does not alter the conclusion given above that the major division in the population of rDNA units is between those that have and those that do not have a segment of DNA inserted at the (β-1)/(β-2) boundary—classes that we choose to call IN$^+$ and IN$^-$, respectively. Along with Wellauer and Dawid (1977), we think that this insertion segment is the major factor in determining the length distribution of the rDNA repeating units detected by electrophoresis of Eco R1 fragments. They have detected some length heterogeneity in the R segment of these units, but it is considerably more limited than that for insertion segments. The agreement between the lengths of R in the long and short units that we observe is consistent with this limitation. The greater heterogeneity in R observed by Pelligrini et al. (1977) is puzzling, but again may be due to a difference in source or isolation procedure.

Models of Transcription

Figure 8 shows the models that we propose for the transcription of the IN$^-$ (A) and IN$^+$ (B) units. The arrangement of the 18S and 28S rDNA in the IN$^-$ units strongly suggests that it is the source of the 8 kb primary transcript that is processed to yield the 18S and 28S rRNAs (Greenberg, 1969; Perry et al., 1970; Jordan et al., 1976). The direction of this transcription is taken from the ultraviolet inactivation studies of J. Carlson and W. Sauerbier (personal communication), which indicate that the 18S sequences are transcribed before the 28S sequences in D. melanogaster cell cultures. This 8 kb primary transcript of the IN$^-$ unit must include all sequences in the 6.9 kb segment of DNA extending from the left end of the 18S rDNA block to the right end of the 28S rDNA block, leaving approximately 1 kb that must be transcribed from the 4.4 kb R segment. We have placed this 1 kb to the left of the 18S rDNA to conform with the transcription map of the rDNA in Xenopus (Wellauer et al., 1976). The remainder of the R segment provides the nontranscribed spacer of the IN$^-$ units (Figure 8A).

Our mapping data, as well as those of Wellauer and Dawid (1977) and of Pellegrini et al. (1977), indicate that the IN$^+$ units are, with the exception of their inserts, very similar if not identical to the IN$^-$ units. Hence we assume that transcription of IN$^+$ units will be initiated as it is in the IN$^-$ units—at a site in the R segment to the left of the 18S rDNA (Figure 8B). The electron microscopic observations of Chooi and Laird on rDNA units in the act of transcription indicate that the IN$^+$ units are indeed transcribed and provide an approximate value for the length of the transcribed region (Chooi, 1976; Laird and Chooi, 1976; Laird et al., 1976; Chooi and Laird, personal communication). They observed that whereas the lengths of the transcribed regions in the tandem arrays are relatively constant and center about a value of approximately 8 kb, the nontranscribed spacers that separate these regions can be grouped into two major length classes with means of approximately 4 and 8 kb. These observations indicate the existence of two major classes of active rDNA units with lengths of about 12 and 16 kb. If we make the reaonable correlation that these two classes equal the major length classes of rDNA units detected by Eco R1 cleavage (Figure 6), then we conclude that the long IN$^+$ units yield a tran-

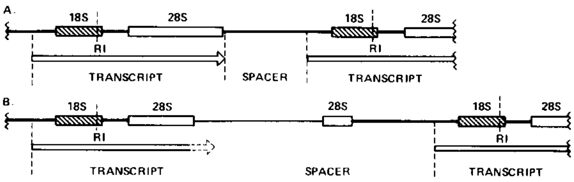


Figure 8. Models for the Transcription of IN$^-$ and IN$^+$ Repeating Units of D. melanogaster rDNA

The dimensions of the IN$^-$ units (A) and IN$^+$ units (B) depicted here are taken, respectively, from the topographies of the short and long units given in Figure 5, where the scale can be taken from the transcript in A, which is 8 kb long (see text).

script that is roughly 8 kb in length. Given that this transcription is initiated as described above, it should end within the insertion segment, as shown in Figure 8B. Hence we postulate that the insertion segment provides a transcription terminator located fairly close (1 ± 1 kb) to the right end of the β-1 section of the 28S rDNA.

The correlation upon which this postulate depends has been confirmed by an examination of the distribution of the long and short units in the nucleolus organizer (NO) of the X and Y chromosomes. Thus Chooi and Laird's electron microscopic observations indicate that the X NO contains both length classes of active units, whereas the long units are absent in the Y NO, and the same distribution has been obtained from an examination of the Eco R1 fragments from each NO (Tartof and Dawid, 1976).

Whether the Y NO contains IN$^+$ units is another question, since IN$^+$ units containing the short insertion segments would not be distinguished by either of the above methods. Hence we do not know whether the tandem arrays of rDNA in the Y NO consist of only IN$^-$ units or of IN$^+$ units randomly interspersed among the IN$^-$ units, as is the case for the X NO (Wellauer and Dawid, 1977). It would appear, however, that the length distribution of insertion segments that was discussed above is characteristic of the NO in the X and not in the Y.

The model for the transcription of IN$^+$ units is qualitative in that it distinguishes only between short and long units. Its generalization to each of the different length classes of IN$^+$ units would depend upon a much more detailed correlation between the two kinds of length distributions than presently exists. It is tempting, however, to adopt the suggestion of Wellauer and Dawid (1977) that insertion segments consist of a tandem array of similar units, each the length of the smallest insert, and to infer that the function of these units is to provide the postulated terminator. In this case, the transcripts of all IN$^+$ units, regardless of the length of the insert, would be alike.

What might be the function of the IN$^+$ transcripts? Although this area of speculation is unrestrained by data, we offer the following testable suggestion. We suppose that the IN$^+$ transcripts are processed to yield the 18S rRNA and an altered form of the 28S rRNA, called 28S* rRNA, and that this 28* rRNA is used at some stage in the life cycle of D. melanogaster – for example, in the nurse cells in which W. Y. Chooi and C. D. Laird observed transcription of the long units. This hypothesis is testable, since the postulated 28S* rRNA is sufficiently defined by its lack of β-2 sequences to be easily identified.

If the IN$^+$ transcript does produce such a 28S* rRNA, or indeed any useful RNA, then it is difficult to imagine why the β-2 segment is maintained in the IN$^+$ units – *unless* a mechanism exists for the interconversion of IN$^-$ and IN$^+$ units. Alternatively, if IN$^+$ units do not yield a useful product, so that the function of the insert is to inactivate the rDNA, then it is difficult to imagine why the IN$^+$ units are retained – *unless* the inserts can be added and/or withdrawn so as to modulate the number of units that can produce the rRNAs according to the demands of different cell states. We find these arguments a strong inducement to determine whether these conversions do indeed occur. Most models for such a conversion will include an additional state, either chromosomal or extrachromosomal, in which the insertion DNA is not integrated within the rDNA. It is therefore of considerable interest that in situ hybridization of ^{3}H-cRNA from Dm103 to polytene chromosomes has revealed sites of hybridization in the centromeric heterochromatin of the autosomes (W. J. Peacock, personal communication) and in many of the bands of the euchromatic arms (D. J. Finnegan, personal communication), in addition to the hybridization previously reported with the nucleolar DNA (Glover et al., 1975). Whether it is the sequences in the insertion segment of Dm103 that are located at these non NO sites is not known, but this possibility certainly stimulates speculation that eucaryotes, like procaryotes, contain specific sequences which migrate within the genome.

Experimental Procedures

Preparative Procedures

pDm103 and pDm102 DNAs were isolated from HB101 hosts as described by Glover et al. (1975), with minor modifications. Total D. melanogaster (Oregon R) chromosomal DNA was isolated from embryonic nuclei as described by Wensink et al. (1974). The 18S and 28S rRNAs were isolated from Schneider's line 2 of D. melanogaster cells as described by Glover et al. (1975) and were a gift from D. M. Glover. ^{32}P-labeled cRNA transcribed from pDm103 with E. coli RNA polymerase was prepared as described by Wensink et al. (1974), except that α-^{32}P-labeled nucleoside triphosphates were substituted for the ^{3}H-labeled substrates.

Enzymes

The sources of Eco R1 and E. coli RNA polymerase are cited in the accompanying paper (Glover and Hogness, 1977). Eco R1 digestions of DNA were carried out at 37°C in 0.01 M MgSO$_4$, 0.1 M Tris-HCl (pH 7.4) and monitored by electrophoresis of the resulting fragments in agarose gels. The fragments to be used for the R loop reaction were precipitated by adding 50 μl of 0.5 M EDTA, 100 μl of 5 M NaCl, and 2.6 ml of 95% ethanol to 1.0 ml of the reaction mixture and, after incubation of this mixture at −20°C for 2 hr, were pelleted by centrifugation (30 min at 40,000 rpm and 0°C in a Beckman SW56 rotor) and suspended in 100 μl of 1 mM EDTA, 10 mM Tris-HCl (pH 7.4). Aspergillus oryzae S1 nuclease (Vogt, 1973) was a gift from T. E. Shenk and was prepared according to the indicated procedure. Pancreatic RNAase was purchased from Worthington Biochemical and was heated for 5 min at 100°C prior to use.

Electron Microscopy

Samples from R loop reaction mixtures described in the figures were prepared for electron microscopy by either of two methods.

Method A: The reaction sample was diluted 50-100 fold in 50% formamide buffer [50% formamide, 0.1 M NaCl, 0.01 M EDTA, 0.1 M Tris-HCl (pH 7.4)] and incubated for 5-10 min at 28°C. To 50 µl of this diluted sample were added 10 µl of 50% formamide, 125 µg cytochrome C per ml, 0.25 M Tris-HCl (pH 8.5), and this mixture was then spread over 10% formamide, 0.01 M Tris-HCl (pH 7.4).

Method B: The reaction sample was diluted 11 fold in the 50% formamide buffer, incubated at 25°C for 10 min, and then diluted 50 fold in 50% formamide, 25 µg cytochrome c per ml, 0.1 M NaCl, 0.01 M EDTA, 0.15 M Tris-HCl (pH 7.4) prior to spreading over 20% formamide, 0.01 M Tris-HCl (pH 7.4).

The short secondary incubation at 25°C or 28°C in 50% formamide is included in both methods to allow reassociation of the DNA strands in regions of Dm103 that although not forming R loops, are nonetheless denatured under the conditions of the R loop reaction—that is, 40°C and 70% formamide buffer (see figure legends). This is particularly true of the L segment (Figure 5), which is frequently found in the denatured state if this secondary incubation is omitted, but appears as duplex DNA when it is included. This behavior is consistent with the low GC content of this segment (approximately 7%) that is estimated from the GC content of the 38S primary transcript and the 18S and 28S rRNAs (31.0%, 40.9%, and 38.1%, respectively; Tartof and Perry, 1970).

After spreading by either method, samples of the film were picked up with parlodin-coated grids, stained with uranyl acetate in 90% ethanol, washed with 90% ethanol, and rotary-shadowed with platinum-palladium. The samples were visualized and photographed in a Phillips 300 electron microscope, and molecular lengths were measured and converted to kb units by comparison to reference DNAs as described in the accompanying paper (Glover and Hogness, 1977). The factor used to convert RNA/DNA duplex lengths to kb was that obtained from reference molecules of pSC101 duplex DNA on the same grid. That the conversion factors for the two kinds of duplexes are equivalent to within ±7% is indicated by a comparison between lengths of regions in Dm103 obtained solely from DNA duplex measurements with those obtained only from RNA/DNA duplexes. For example, the region between the left ends of the α and β-1 loops can be obtained by subtracting L (1.44 ± 0.019 kb; N = 20; Figure 5A) from the length of the DNA duplex extending from the left end of Dm103 to the left end of the β-1 loop in molecules that have no α loop (3.28 ± 0.043 kb; N = 13). The resulting value of 1.84 ± 0.047 kb is to be compared with a value of 1.81 ± 0.080 kb for the same region but obtained from the differences in lengths of the RNA/DNA duplexes in the combined α + (β-1) loop and in the β-1 loop (see Results) when using the DNA duplex conversion factor. This comparison indicates that the ratio of the actual conversion factor for RNA/DNA duplexes to that for DNA duplexes is 1.02 ± 0.05. A similar comparison for the region between the left and right ends of the α + (β-1) combined loops yields a ratio of 0.99 ± 0.05.

Enrichment for the Long and Short Repeating Units by Zone Sedimentation

1 ml of Eco R1-digested embryonic DNA (80 µg/ml) was layered on top of a 10-30% sucrose gradient in 0.2 M NaCl, 0.001 M EDTA, 0.01 M Tris-HCl (pH 7.4) (16 ml) and centrifuged in a Beckman SW27.1 rotor at 27,000 rpm and 20°C for 11 hr. An Eco R1 digest of ^{32}P-labeled cDm103 hybrid DNA (Grunstein and Hogness, 1975) was identically centrifuged to provide markers for the identification of the peak fractions of the long (17 kb) and short (11 kb) units. Since Eco R1 cleavage of cDm103 yields a 17 kb (Dm103) and a 6.5 kb (Col El) fragment, the fraction containing the long units was directly identified by the longer marker, and that containing the shorter units was determined by interpolation between the markers, according to the equation $d = KM^{\alpha}$ (where d and M are the distance sedimented and molecular length, respectively, and K and α are constants; see Egan and Hogness, 1972). Each gradient was divided into 39 equal fractions, and the 20th and 23rd fraction from the bottom were taken for the long and short fragments, respectively. The DNA in these fractions was precipitated with ethanol prior to use for R loop analysis.

Containment Conditions

All plasmids containing Dm segments were propagated in EK1 host-vector systems under P2 physical containment, as defined and recommended for this class of experiments by Guidelines for Research Involving Recombinant DNA Molecules issued in June, 1976 by the National Institutes of Health, Bethesda, Maryland.

Acknowledgments

We thank Dr. D. M. Glover for his generous contributions to certain experiments reported here and for many helpful discussions. We are indebted to Drs. Laird, Carlson, Sauerbier, Wellauer, Dawid, Pellegrini, Manning, and Davidson for unpublished information. This work was supported by research grants from the NSF and the NIH. R.L.W. was a Jane Coffin Child fellow.

Received September 13, 1976; revised October 22, 1976

References

Birnstiel, M. L., Sells, B. H., and Purdom, I. F. (1972). Kinetic complexity of RNA molecules. J. Mol. Biol. *63*, 21-39.

Chooi, W. Y. (1976). RNA transcription and ribosomal protein assembly in D. melanogaster. In Handbook of Genetics, 5, R. C. King, ed. (New York: Plenum Press), pp. 219-265.

Davis, R. W., Simon, M., and Davidson, N. (1971). Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. In Methods in Enzymology, *21D*, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 413-428.

Egan, J. B., and Hogness, D. S. (1972). The topography of lambda DNA: isolation of ordered fragments and the physical mapping of point mutations. J. Mol. Biol. *71*, 363-381.

Glover, D. M., and Hogness, D. S. (1977). A novel arrangement of the 18S and 28S sequences in a repeating unit of Drosophila melanogaster rDNA. Cell *10*, 167-176.

Glover, D. M., White, R. L., Finnegan, D. J., and Hogness, D. S. (1975). Characterization of six cloned DNAs from Drosophila melanogaster, including one that contains the genes for rRNA. Cell *5*, 149-157.

Greenberg, J. R. (1969). Synthesis and Properties of Ribosomal RNA in Drosophila. J. Mol. Biol. *46*, 85-98.

Grunstein, M., and Hogness, D. S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Nat. Acad. Sci. USA *72*, 3961-3965.

Jordan, B. R., Jourdan, R. and Jacq, B. (1976). Late steps in the maturation of Drosophila 26S ribosomal RNA: generation of 5.8S and 2S RNAs by cleavages occurring in the cytoplasm. J. Mol. Biol. *101*, 85-105.

Laird, C. D., and Chooi, W. Y. (1976). Morphology of transcription units in Drosophila melanogaster. Chromosoma *58*, 193-218.

Laird, C. D., Wilkinson, L. E., Foe, V. E., and Chooi, W. Y. (1976). Analysis of chromatin-associated fiber arrays. Chromosoma *58*, 169-192.

Lee, C. S., Davis, R. W., and Davidson, N. (1970). A physical study by electron microscopy of the terminally repetitious circularly permuted DNA from the coliphage particles of Escherichia coli 15. J. Mol. Biol. *48*, 1-22.

Loening, U. E. (1968). Molecular weights of ribosomal RNA in relation to evolution. J. Mol. Biol. *38*, 355-365.

Meyer, G. F., and Hennig, W. (1974). The nucleolus in primary spermatocytes of Drosophila hydei. Chromosoma *46*, 121-144.

Pellegrini, M., Manning, J., and Davidson, N. (1977). Sequence arrangement of the rDNA of Drosophila melanogaster. Cell *10*, 000-000.

Perry, R. P., Cheng, T.-Y., Freed, J. J., Greenberg, J. R., Kelley, D. E., and Tartof, K. D. (1970). Evolution of the transcription unit of ribosomal RNA. Proc. Nat. Acad. Sci. USA *65,* 609-616.

Shine, J., and Dalgarno, L. (1973). Occurrence of heat-dissociable ribosomal RNA in insects: the presence of three polynucleotide chains in 26S RNA from cultured Aedes aegypti cells. J. Mol. Biol. *75,* 57-72.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98,* 503-517.

Tartof, K. D., and Perry, R. P. (1970). The 5S RNA genes of Drosophila melanogaster. J. Mol. Biol. *51,* 171-183.

Tartof, K. D., and Dawid, I. B. (1976). Similarities and differences in the structure of the X and Y chromosome rRNA genes in Drosophila. Nature *263,* 27-30.

Thomas, M., White, R. L., and Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc. Nat. Acad. Sci. USA *73,* 2294-2298.

Vogt, V. M. (1973). Purification and further properties of single-strand specific nuclease from Aspergillus oryzae. Eur. J. Biochem. *33,* 192-200.

Wang, J. C. (1974). Interactions between twisted DNAs and enzymes: the effects of superhelical turns. J. Mol. Biol. *87,* 797-816.

Wellauer, P. K., and Dawid, I. B. (1977). The structural organization of ribosomal DNA in Drosophila melanogaster. Cell *10,* 000-000.

Wellauer, P. K., Dawid, I. B., Brown, D. D., and Reeder, R. H. (1976). The molecular basis for length heterogeneity in ribosomal DNA from Xenopus laevis. J. Mol. Biol., in press.

Wensink, P. C., Finnegan, D. J., Donelson, J. E., and Hogness, D. S. (1974). A system for mapping DNA sequences in the chromosomes of Drosophila melanogaster. Cell *3,* 315-325.