The Chromosomal Arrangement of Coding Sequences In a Family of Repeated Genes

G. M. RUBIN,[1]
D. J. FINNEGAN[2] AND
D. S. HOGNESS

Department of Biochemistry
Stanford University School of Medicine
Stanford, California

We are interested in studying the chromosomal arrangement of DNA sequences that code for mRNAs in *Drosophila melanogaster,* and in analyzing adjacent sequences that may control their expression. In order to isolate individual segments coding for particular mRNAs, we have constructed a set of hybrid DNA molecules by joining [with the aid of terminal transferase (*1*)] the bacterial plasmid ColE1 to sheared fragments of *D. melanogaster* embryonic DNA. From among these, we have identified a single DNA segment that contains sequences homologous to approximately 1% of the mass of cytoplasmic poly(A) containing RNA from *D. melanogaster* tissue culture cells. Some of the properties of this hybrid, which we have called cDm412, are shown in Fig. 1 (*2, 3*).

The mRNA species complementary to cDm412 is 6000–7000 nucleotides long as determined by polyacrylamide gel electrophoresis in 96% formamide. The sequences on cDm412 homologous to this message are confined to the restriction fragments A, B, C, D, E and F (Fig. 1). These are internal fragments and span a distance 9500 nucleotides indicating that cDm412 can carry only one copy of this mRNA sequence. The poly(A)-containing end of the RNA lies in fragment A, suggesting that transcription is from right to left on the map (Fig. 1; *3*). In addition to sequences complementary to an abundant mRNA, cDm412 also contains sequences representative of several families of moderately repetitive DNA sequences. Studies on the distribution and interrelationships of these moderately repetitive sequences are described elsewhere(*2*).

[1] Present address: Sidney Farber Cancer Center, Harvard Medical School, 35 Binney Street, Boston, Massachusetts 02115.

[2] Present address: Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland.

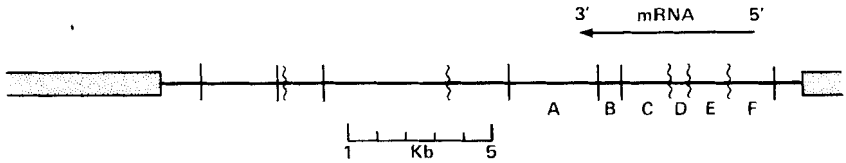

FIG. 1. Physical map of cDm412. The thin horizontal line represents *Drosophila* DNA. The thick line represents DNA of the plasmid vector ColE1. The circular map has been opened at a Sma 1 restriction enzyme cleavage site within the ColE1 DNA to produce the linear map shown. The vertical lines represent the cleavage sites for the restriction enzymes EcoR1 ($\S$) and HindIII (|). A scale in kilobases (1000 nucleotides = 1 kb) is shown. The approximate location of the mRNA sequences is shown above the map.

We have determined how the sequences of cDm412 are arranged within the *D. melanogaster* genome by *in situ* hybridization of polytene salivary gland chromosomes with [^{3}H]RNA complementary to cDm412. About 70 sites on the chromosome arms as well as the chromocenter were labeled. The cDm412 sequences homologous to mRNA must lie at one or more of these sites. In fact, about 30 sites on the chromosome arms were labeled after *in situ* hybridization using a probe made from fragment E, which contains only mRNA sequences. A similar pattern of labeling is seen after *in situ* hybridization using [^{3}H]RNA complementary to the HindIII endonuclease fragment containing sequences C, D, E and F (Fig. 2). The entire sequence coding for the mRNA may be present at each of these sites, or alternatively the labeling at some of them might be due to homology to fragment E alone.

Two lines of evidence suggest that the entire mRNA sequence is represented at most, perhaps at all of these sites. The first comes from *in situ* hybridization of [^{3}H]cRNA to fragment A. If the entire mRNA is present at each site labeled by fragment E, all these sites should also be labeled by fragment A. With the help of M. Young, we have mapped the 10 sites on the X chromosome labeled by fragment E and have compared them with those sites labeled by fragment A. Indeed, all sites on the X chromosome labeled by fragment E are also labeled by fragment A.

In order to examine in more detail the sequences present at several of the 30 or so chromosomal sites, we have screened several thousand independently cloned hybrid plasmids for those that contain sequences present in fragment E. The screen was carried out by the colony hybridization method of Grunstein and Hogness (*4*), using [^{32}P]cRNA to fragment E as the probe. Four of the desired hybrids were isolated, and two of these, cDm454 and cDm468, are compared with cDm412 in Fig. 3.

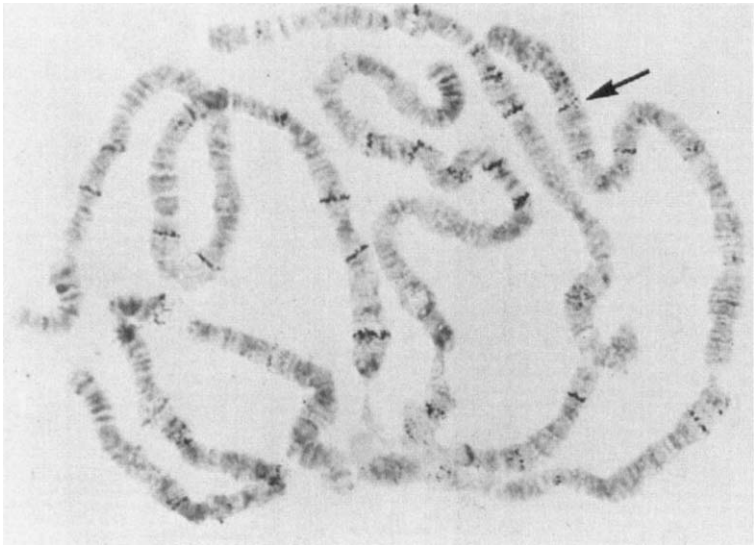

FIG. 2. *In situ* hybridization of *Drosophila melanogaster* polytene chromosomes with [^{3}H]cRNA to the HindIII endonuclease fragment containing sequences C, D, E and F (Fig. 1). *In situ* hybridization was carried out as described previously (*1*). The arrow indicates labeling of the 3C2-7 region of the X-chromosome.

cDm454 and cDm468 both yield fragments identical in size to fragments B, C, D and E of cDm412 after digestion with the restriction enzymes HindIII and EcoR1, and in each case these fragments show sequence homology to mRNA. By contrast, the two fragments that contain DNA complementary to the ends of the mRNA, as well as adjacent non-mRNA sequences (for example fragments A and F of cDm412) are different in each case. We conclude that: (i) the three cloned segments are derived from different chromosomal sites; (ii) these sites contain the same or very similar mRNA sequences; and (iii) the different sites contain different sequences adjacent or close to the ends of these mRNA sequences.

Taken together, this evidence strongly suggests that the structural gene carried by cDm412 is present at each of approximately 30 sites in the genome. How faithful is this repetition? We have not yet obtained a quantitative measure of the degree of mismatch among the mRNA sequences at different sites. The mRNA regions could exhibit small differences due either to third-position variation in codon sequences that are otherwise identical, or to slight perturbations in the amino-acid sequences coded from each mRNA. However, the observation that the EcoR1 and HindIII cleavage sites are identically distributed within the mRNA

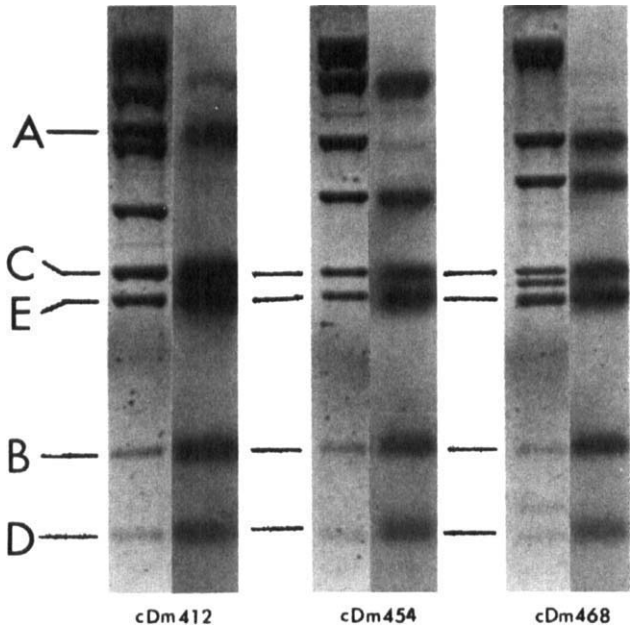

FIG. 3. A comparison of three independent hybrid plasmids that have sequence homology to fragment E. The left-hand portion of each panel shows the fragments generated by digesting the plasmids with both of the restriction endonucleases EcoR1 and HindIII. The fragments were separated by electrophoresis through 1.4% agarose gels containing 0.09 M Tris-borate, 3 mM EDTA, 2 μg/ml ethidium bromide, pH 8.4. The gels were photographed under UV illumination.

The DNA in each gel was transferred to a nitrocellulose filter by the procedure of Southern (5). In order to determine which restriction enzyme fragments contained sequences complementary to mRNA, ^{32}P-labeled poly(A)-containing cytoplasmic RNA from *D. melanogaster* tissue culture cells (6) was hybridized to each filter. Hybridization reactions were carried out for 16 hours at 43°C in 0.1 M sodium phosphate, 0.6 M NaCl, 0.06 M sodium citrate, 50% formamide, pH 7, containing 200 μg of poly(A) per milliliter. The right-hand portion of each panel shows the autoradiograph of the filter after unhybridized RNA had been removed by RNase treatment and successive washes in hybridization buffer and then with 0.3 M NaCl, 0.03 M sodium citrate. In the digest of cDm412, fragment F comigrates with fragment C; these have been distinguished by digestion with other restriction enzymes (3).

sequences at three chromosomal sites indicates that in this small sample there is little if any variation in the number of nucleotides in each of the regions defined by these cleavages. The following experiment indicates that the region corresponding to fragment C in cDm412 is likewise invariant for most, if not all, of the chromosomal sites.

Total *D. melanogaster* DNA was digested by EcoR1 and HindIII restriction endonucleases and the resulting fragments fractionated accord-

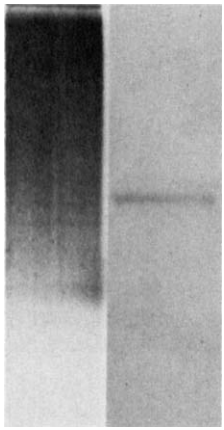

FIG. 4. Homogeneity of fragment C sequences within the genome. The left-hand panel shows the products of a combined HindIII,EcoR1 digest of *Drosophila melanogaster* embryo DNA after separation by electrophoresis as described in the legend to Fig. 3. The DNA was transferred to nitrocellulose as before and the nitrocellulose filter was then treated by the procedure of Denhardt (7). DNA from fragment C was labeled with ^{32}P to a specific activity of ca. 5×10^7 cpm/μg by the "nick translation" reaction of DNA polymerase I (8) and hybridized to the filter by a modification of the Denhardt (7) procedure. The hybridization was carried out at 65°C at a salt concentration of 0.75 M NaCl, 0.075 M sodium citrate. After 36 hours of incubation, the filters were washed exhaustively with 0.3 M NaCl, 0.03 M sodium citrate at 65°C. A radioautograph of the filter is shown in the right-hand panel.

ing to length by electrophoresis in a 1.4% agarose gel. These fragments were then denatured and transfered to a nitrocellulose filter (5), and those that contain sequences homologous to fragment C were assayed by hybridization with ^{32}P-labeled fragment C DNA. Figure 4 shows that more than 95% of this hybridization was restricted to a single class of fragments identical in length to fragment C.

We wish to know which of these repeated genes are transcribed in which cell types, and to map the sequences that are transcribed for at least some of these genes. Restriction fragments adjacent to the mRNA sequence should be useful probes for these purposes. We know that sequences adjacent or very close to some copies of the mRNA sequence are different. If this is generally true, and if these sequences are transcribed, then we should be able to solve these problems by the use of such probes. One of the sites labeled by fragment E is within the genetically well defined 3C2-7 region of the X chromosome (9), and we anticipate that this will allow a combined genetic and biochemical attack on the genes in this region.

References

1. P. C. Wensink, D. J. Finnegan, J. E. Donelson and D. S. Hogness, *Cell* 3, 315 (1974).
2. D. J. Finnegan, G. M. Rubin, D. J. Bower and D. S. Hogness, (1976). In preparation.
3. G. M. Rubin, D. J. Finnegan and D. S. Hogness, (1976). In preparation.
4. M. Grunstein and D. S. Hogness, *PNAS* **72**, 3961 (1975).
5. E. M. Southern, *JMB* **98**, 503 (1975).
6. G. M. Rubin and D. S. Hogness, *Cell* **6**, 207 (1975).
7. D. T. Denhardt, *BBRC* **23**, 641 (1966).
8. F. H. Schachat and D. S. Hogness, *CSHSQB* **38**, 371 (1973).
9. G. Lefevre and M. M. Green, *Chromosoma* **36**, 391 (1972).