Different Regions of a Complex Satellite DNA Vary in Size and Sequence of the Repeating Unit

MARIAN CARLSON† AND DOUGLAS BRUTLAG

Department of Biochemistry
Stanford University School of Medicine
Stanford, Calif. 94305, U.S.A.

(Received 29 June 1979)

The 1.688 g/cm³ satellite DNA of Drosophila melanogaster is composed primarily of 359 base-pair units repeated in tandem. Most of these units contain a single cleavage site for both HaeIII and HinfI restriction endonucleases; however, some units lack one or both sites. Previously we had shown that the distribution of HaeIII and HinfI endonuclease sites varies widely between different regions of 1.688 g/cm³ satellite DNA; for example, some regions contain HaeIII sites in every unit and other regions (>10,000 base-pairs) contain no HaeIII sites (Carlson & Brutlag, 1977). We have now cloned molecules of 1.688 g/cm³ satellite DNA which lack HaeIII sites and have shown that the absence of sites is caused by sequence variation rather than base modification. This result indicates that regions of 1.688 g/cm³ satellite DNA with different distributions of restriction sites differ in the sequence of their repeating units. We also show that a large fraction of the satellite DNA which is not cleaved by HaeIII endonuclease still contains HinfI endonuclease sites (and AluI sites) spaced about 359 base-pairs apart. However, one cloned segment lacking HaeIII sites was found to contain 33 tandem copies of a novel 254 base-pair unit. Sequence analysis showed that this 254 base-pair unit is homologous to the 359 repeat except for a 98 base-pair deletion. These data suggest that both units have evolved from a common ancestor and that each has subsequently become amplified into separate tandem arrays.

1. Introduction

The genome of Drosophila melanogaster contains several highly repeated satellite DNAs, which comprise about 20% of the genome and are primarily located within centromeric heterochromatin. Analysis of the sequence organization of these satellite DNAs is essential for understanding their evolution, and will perhaps provide clues to their function. We have studied the 1.688 g/cm³ satellite (1.688 satellite) DNA, which is mainly composed of tandem repeats of a 359 base-pair unit (Hsieh & Brutlag, 1979).

Previous studies have shown that although many 359 base-pair units contain a recognition site for both HaeIII and HinfI endonucleases, some units lack one or both sites (Manteuil et al., 1975; Shen et al., 1976). Digestion of 1.688 satellite DNA with either enzyme generates monomers, dimers, trimers, and so on, of the 359 base-pair unit, as well as very large fragments (>10⁶ base-pairs) resistant to cleavage.

In a previous report we presented evidence that 1.688 satellite DNA includes regions

† Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, Mass. 02139, U.S.A.
with different distributions of HaeIII endonuclease sites. Some regions contain few or no HaeIII endonuclease cleavage sites, while most regions contain sites almost every 359 base-pairs (Carlson & Brutlag, 1977). Because the absence of a recognition site could result from either sequence alteration or base modification of the recognition sequence, it was unclear whether these different regions of 1-688 satellite DNA differ in nucleotide sequence or in degree of base modification.

In this work we have investigated the absence of HaeIII endonuclease sites in both the dimer fragments and long HaeIII-resistant fragments produced from 1-688 satellite DNA by HaeIII endonuclease cleavage. Using molecular cloning, we have shown that nucleotide sequence, not base modification, is responsible for the lack of HaeIII sites. Hence, regions of 1-688 satellite DNA with different distributions of HaeIII endonuclease sites differ in sequence. We have characterized 1-688 satellite DNA regions which lack HaeIII sites to determine their relationship to the bulk of the 1-688 satellite DNA, which contains both HaeIII and HinFI endonuclease sites almost every 359 base-pairs. Much of the HaeIII-resistant DNA was shown to be composed of repeated units approximately 359 base-pairs in size and to contain regularly spaced HinFI endonuclease sites. However, analysis of cloned segments of HaeIII-resistant satellite DNA led to the discovery of a novel 1-688 satellite DNA component with repeating units of 254 base-pairs. Sequence analysis of the 254 base-pair unit showed that it is homologous to all but 98 base-pairs of the 359 base-pair unit. These data suggest that both repeat units evolved from a common ancestor and that each was subsequently amplified to generate separate tandem arrays of the 359 and 254 base-pair repeats.

2. Materials and Methods

(a) Preparation of DNA

The 1-688 satellite DNA was purified from D. melanogaster embryonic nuclei (Carlson & Brutlag, 1977). Analytical centrifugation showed that 36% of the recovered DNA had a buoyant density of 1-6885 g/cm³, and the remaining 64% formed a broad peak with a density of 1-689 g/cm³. The DNA sample used for the experiment shown in Fig. 1 was further purified by centrifugation in cesium formate and is the same preparation described before (Carlson & Brutlag, 1977). Dimer fragments and long HaeIII-resistant fragments were purified from 1-688 satellite DNA by digestion to completion with HaeIII endonuclease. The products were then fractionated by electrophoresis in a 1% agarose gel. The dimer fragments and the HaeIII-resistant fragments migrating at the limiting mobility were eluted from the gel, as described below. The HaeIII-resistant fragments were estimated to be larger than 10 × 10² base-pairs by gel electrophoretic comparison with size standards. Procedures for isolation of closed circular plasmid DNAs have been described (Carlson & Brutlag, 1977).

(b) Enzymes and reactions

Terminal deoxynucleotidyl transferase (Chang & Bollum, 1971), HaeIII endonuclease (Roberts et al., 1975) EcoRI endonuclease (Modrich & Zabel, 1976) and polynucleotid kinase (Richardson, 1965) were gifts of R. Ratliff, D. Charney, P. Modrich and J. Chien, respectively. E. coli polymerase I was purified as described by Jovin et al. (1969) and HinFI endonuclease was prepared by BioGel A-0-5 M, phosphocellulose and DEAE-cellulose chromatography, as described by R. J. Roberts (personal communication). AluI and HaeIII endonucleases were purchased from New England Biolabs. Bacterial alkaline phosphatase was purchased from Worthington and further purified by DEAE-cellulose chromatography by M. Goldberg. Reaction conditions for restriction endonuclease digestion have been described (Carlson & Brutlag, 1977).
SEQUENCE VARIATIONS IN SATELLITE DNA

(c) Bacteria

E. coli strains HB101 (recA, hsr, hsm) (Boyer & Roulland-Dussoix, 1969) and C600 (pSC101) (Cohen et al., 1973) were obtained from D. S. Hogness.

(d) Construction of hybrid DNA and transformation of bacteria

Hybrid molecules were constructed by the poly(dA)·poly(dT) joining methods of Brutlag et al. (1977). For preparation of pSC105–dimer hybrid molecules, purified 1-688 satellite dimer fragments were extended with poly(dA)$_{60}$ using terminal deoxynucleotidyltransferase. D. Kemp generously provided BanHI endonuclease-cleaved pSC105 DNA (Cohen et al., 1973) to which poly(dT)$_{60}$ had been added, and these 2 DNA samples were allowed to anneal. Similarly, purified HaeIII-resistant fragments and EcoRI-cleaved pSC101 were extended with poly(dT)$_{60}$ and poly(dA)$_{60}$, respectively. Transformation was effected as described before (Carlson & Brutlag, 1977). Plates containing kanamycin (33 µg/ml) or tetracycline (15 µg/ml) were used to select bacteria transformed with pSC105 or pSC101 DNA, respectively, and transformed colonies were tested for the presence of 1-688 satellite DNA according to a modification of the colony hybridization technique of Grunstein & Hogness (1975) (J. Lis and L. Prestidge, personal communication). The hybridization probe used in this procedure was prepared as follows. A recombinant plasmid containing 1-688 satellite DNA inserted into a vector lacking homology to pSC101 or pSC105 was constructed by joining poly(dA)$_{60}$-terminated 1-688 satellite DNA to poly(dT)-terminated ColE1 DNA (a gift from G. Rubin). Selection for transformed colonies was accomplished by addition of colicin E1 (from D. Finnegan) to transform bacteria before plating. The plasmid cDm688-52, containing an inserted segment of 1-688 satellite DNA, was recovered. $^{32}$P-labeled cDm688-52 DNA (5 x 10$^7$ cts/min per µg) was prepared by nick-translation with E. coli DNA polymerase I, using $^{32}$PdGTP (200 Ci/mmol; Amersham) as labeled substrate, according to the method of Rigby et al. (1977). Work involving recombinant plasmids was carried out under the P2 level of physical containment and EK1 level of biological containment, as specified in the National Institutes of Health Guidelines on Recombinant DNA Research (June 1976).

(e) Gel electrophoresis and elution of DNA from gels

Agarose gel electrophoresis has been described (Carlson & Brutlag, 1977). Polyacrylamide gels were prepared in TBE buffer (89 mm-Tris·OH, 89 mm-boric acid, 2-5 mm-EDTA, pH 8-3; Peacock & Dingman, 1968) from a stock of 20% acrylamide, 1% N,N'-methylenebisacrylamide. Analysis of restriction enzyme cleavage patterns was carried out on 3% to 7% linear polyacrylamide gradient slab gels (14 cm x 14 cm x 0-16 cm) by electrophoresis at 100 V for 3 h. For preparation of DNA fragments, polyacrylamide slab gels (23 cm x 14 cm x 0-16 cm, or 23 cm x 14 cm x 0-3 cm) were used, and electrophoresis was at 60 V for 12 h. DNA fragments were recovered from gels by electrophoretic elution in either TBE buffer or 0-05 x TBE buffer at 50 V overnight. DNA was electrophoresed into dialysis tubing, and the polarity was reversed for 1 min prior to recovery of DNA from the bag. Residual ethidium bromide was sometimes removed by chromatography on Dowex AG50W-X8. The DNA was then precipitated with ethanol and resuspended in TE buffer (10 mm-Tris·HCl, 1 mm-EDTA, pH 8).

(f) Hybridization to DNA transferred from a gel to nitrocellulose

The pattern of DNA fragments was transferred from gels to nitrocellulose sheets according to Southern (1975). Sheets were pre-incubated several hours at 65°C, except where noted, in 0-02% bovine serum albumin, 0-02% Ficoll (Pharmacia), 0-02% polyvinyl pyrrolidone, 5 x SSCP (0-6 m-NaCl, 0-075 m-sodium citrate, 0-1 m-sodium phosphate, pH 7) and 0-1% sodium dodecyl sulfate (Denhardt, 1966). Hybridization with $^{32}$P-labeled cDm688-52, prepared as above, was carried out for 12 to 48 h under the same conditions as the pre-incubation, except that carrier salmon sperm DNA (100 µg/ml) and poly(A) (100 µg/ml) were also included. The nitrocellulose sheets were washed with shaking 3 times in pre-incubation buffer at the hybridization temperature and then twice in 2 x SSC (SSC is 0-15 m-NaCl, 0-015 m-sodium citrate, pH 7) at room temperature; each wash was for 30 min.
DNA was prepared for electron microscopy by the basic protein film technique with either 40% or 70% formamide (Davis et al., 1971). Heteroduplexes were formed by the procedure of Davis et al. (1971), except that renaturation was carried out in either 70% formamide at 25°C or 45% formamide at 37°C.

(h) DNA sequence analysis

For determination of nucleotide sequences adjacent to AluI endonuclease cleavage sites, pDm888-169 (20 μg) was digested with AluI endonuclease, then precipitated with ethanol. The resulting fragments were dephosphorylated by incubation with bacterial alkaline phosphatase in 20 mM-Tris·HCl (pH 8) at 60°C for 2 h, and the enzyme was then removed with 4 phenol extractions. After ether extraction and precipitation with ethanol, the DNA was resuspended in 50 μl of 10 mM-Tris·HCl (pH 8), 50 mM-KCl, 0.1 mM-EDTA, 5’-phosphorylation was then carried out by addition of 5 μl of 0.1 M-MgCl₂, 80 mM-dithiothreitol and 50% glycerol; 200 pmol of [γ-32P]ATP (~1000 Ci/mmol, prepared according to Maxam & Gilbert (1977)); and 2 μl polynucleotide kinase (1-1 units/ml). The reaction was incubated at 37°C for 30 min and terminated by addition of an equal vol. of 4 M-ammonium acetate. The DNA was then precipitated with ethanol twice in the presence of carrier transfer RNA (20 μg). The labeled fragments were separated by electrophoresis on a 2% agarose gel, and the 254 base-pair monomer fragments were eluted from the gel as described above. These fragments were next digested with HindIII endonuclease. The 254 base-pair fragments containing a HindIII site gave rise to 2 cleavage products (88 and 166 base-pairs), which were separated from the uncleaved 254 base-pair fragments by electrophoresis on a 5% polyacrylamide gel and eluted from the gel.

Fragments suitable for analysis of sequences adjacent to HindIII endonuclease sites were prepared as follows. pDm888-169 DNA (150 μg) was digested with HindIII endonuclease, and the resulting fragments were fractionated by 5% polyacrylamide gel electrophoresis. The 254 base-pair fragment of satellite DNA co-migrates with a fragment of pSC101 DNA, and both fragments were eluted from the gel. The DNA was then extracted twice with n-butanol (to remove the ethidium bromide used in staining DNA in the gel), and loaded on a small DEAE-cellulose (DE52) column (0·2 ml) equilibrated with 0·1 M-Tris·HCl (pH 7·5). After washing the column with this buffer, the DNA was eluted with 0·7 M-NaCl in 0·1 M-Tris·HCl (pH 7·5). Closed circular carrier DNA (12 μg) was added, and the DNA was precipitated with ethanol. Dephosphorylation and 5’-phosphorylation reactions were carried out essentially as described above, using 100 pmol [γ-32P]ATP (2500 Ci/mmol; ICN). The labeled fragments were then digested with AluI endonuclease and electrophoresed on a 5% polyacrylamide gel.

All 4 fragments (2 satellite DNA fragments and 2 vector DNA fragments) were recovered, chromatographed on DEAE-cellulose as described above, and ethanol precipitated with carrier tRNA (25 μg). The satellite DNA fragments (88 and 166 base-pairs) were identified on the basis of size, but all 4 fragments were sequenced to confirm the identification.

Chemical reactions and electrophoresis on sequencing gels were carried out according to Maxam & Gilbert (1977). In addition to 20% and 25% polyacrylamide gels, a 12% acrylamide/8-6% bisacrylamide gel in 7 M-urea and 100 mM-Tris-borate (pH 8·3), 1 mM-EDTA was also employed.

3. Results

(a) Sequence variation rather than base modification accounts for absence of HaeIII endonuclease recognition sites

Most of the 359 base-pair repeating units of 1-688 satellite DNA contain a recognition site for HaeIII endonuclease, but many units lack such a site; consequently, digestion of 1-688 satellite DNA with HaeIII endonuclease produces not only 359 base-pair monomers, but also dimers, trimers, and higher oligomers, including large fragments which migrate with limiting mobility on a gel (Fig. 1). The absence of
Fig. 1. Fragments generated by *Hae*III and *Hin*II endonuclease digestion of 1.688 satellite DNA. The 1.688 satellite DNA, purified as described in Materials and Methods, was digested to completion with *Hae*III endonuclease or *Hin*II endonuclease, and the fragments were electrophoresed in a 1% agarose gel. bp, base-pairs.
HaeIII cleavage sites in the fragments larger than monomer could result from either sequence variation or base modification of HaeIII recognition sequences. To distinguish between these possibilities, we cloned dimer fragments and fragments migrating with limiting mobility on a gel and then examined the cloned DNA for the presence of HaeIII sites. Since DNA propagated in host-specific modification-defective E. coli exhibits no base modification of HaeIII sites (Carlson & Brutlag, 1977), the presence or absence of HaeIII endonuclease sites in cloned DNA is determined only by the nucleotide sequence.

Dimer fragments were purified from HaeIII endonuclease-cleaved 1-688 satellite DNA and inserted in vitro into the BamHI site of the kanamycin resistance plasmid vector pSC105 by the poly(dA)·poly(dT) joining method, as described in Materials and Methods. The resulting molecules were used to transform E. coli HB101 to kanamycin resistance, and transformed colonies were then screened for the presence of 1-688 satellite DNA by colony hybridization (Grunstein & Hogness, 1975). Colonies containing 1-688 satellite DNA carried hybrid plasmids with inserted satellite dimer fragments, called pSC105–dimer plasmids.

Plasmid DNA was prepared from 18 clones containing 1-688 dimer units, which were recovered from ten independent transformations. The expected structure of these pSC105–dimer plasmids is shown in Figure 2. To confirm that the hybrid

![Diagram of pSC105–dimer plasmid structure](image)

**Fig. 2.** Expected structure of pSC105–dimer plasmids. The diagram shows schematically (not to scale) the expected structure of a pSC105–dimer hybrid plasmid. The single line represents the inserted dimer fragment, the double line represents the pSC105 vector, and the black regions at the junctions between dimer and pSC105 represent poly(dA)·poly(dT). In each DNA strand the dimer is flanked by poly(dA) at one junction and poly(dT) at the other. The hatched areas along the pSC105 segment represent an inverted repeat in the kanamycin resistance element (Guyer et al., 1976).

plasmids have this structure, four plasmid DNA preparations were cleaved with EcoRI endonuclease, and the resulting fragments were denatured, allowed to anneal in a first-order reaction as described in the legend to Figure 3, and mounted for electron microscopy. The poly(dA) and poly(dT) segments were expected to form a short stem and dimer-sized loop, and the two inverted repeat sequences in the kanamycin resistance element (Guyer et al., 1976) were expected to form a long stem and small loop. Figure 3 shows that such structures were observed in the electron microscope, indicating that the four pSC105–dimer plasmids tested have the structure shown in Figure 2.

Each pSC105–dimer DNA sample was then digested with HaeIII endonuclease to
determine whether the inserted dimer fragment could be cleaved. If the inserted dimer had no *HaeIII* site, digestion would generate, in addition to fragments derived from the pSC105 vector, a single novel fragment containing the inserted dimer flanked by poly(dA)-poly(dT) segments and adjacent vector DNA. On the other hand, if the inserted dimer were cleaved by *HaeIII* endonuclease, two new fragments would be generated. Each of 18 pSC105–dimer plasmids gave rise to only one new
fragment which was, as expected, slightly larger than a 1·688 dimer fragment isolated from satellite DNA. Figure 4 shows the HaeIII digestion products of eight representative plasmids; the size variation of the new fragments containing the inserted dimers is presumably due to differences in length of the poly(dA)·poly(dT). Similarly, the presence of two new fragments, both larger than a dimer fragment, in slot four probably indicates heterogeneity in the length of the poly(dA)·poly(dT) segments within this preparation of plasmid DNA (Brutlag et al., 1977).

![Image of gel electrophoresis](image)

**Fig. 4.** Cloned dimer fragments lack HaeIII endonuclease sites.

Eighteen pSC105–dimer hybrid plasmids were digested with HaeIII endonuclease, and the digestion products were electrophoresed in 1·6% agarose gels. The gel shown here displays the fragments generated from 8 representative pSC105–dimer plasmids and the vector pSC105. The fragments containing inserted dimers are indicated by a bracket. The fragments generated from partially purified 1·688 satellite DNA by HaeIII digestion are also shown, and the monomer and dimer fragments are indicated by arrows.
In a similar experiment, we also investigated whether sequence variation or base modification is responsible for the absence of HaeIII recognition sites in the long satellite DNA fragments which migrate at the limiting mobility on a gel after HaeIII digestion. These HaeIII-resistant fragments were purified as described in Materials and Methods and inserted into the EcoRI endonuclease cleavage site of the tetracycline resistance plasmid vector pSC101 by the poly(dA)·poly(dT) joining method. Fourteen clones containing 1-688 satellite sequences were recovered from seven independent transformations, and hybrid plasmid DNA was isolated. When cleaved with HaeIII endonuclease, 13 of the 14 plasmids produced one novel fragment, in addition to the fragments expected from the pSC101 vector (Fig. 5). One hybrid plasmid (Fig. 5, slot 3) gave rise to two new fragments; the presence of one HaeIII cleavage site indicates either that the D. melanogaster DNA fragment from which this hybrid plasmid was derived contained a single modified HaeIII site or that the initial HaeIII endonuclease digestion was incomplete.

Of the 14 hybrid plasmids, six carried inserted segments larger than $3.5 \times 10^3$ base-pairs (see the legend to Fig. 5). Most of the inserted DNA in the six largest hybrid plasmids was shown to be homologous to 1-688 satellite DNA by its ability to hybridize to cloned satellite DNA (see Fig. 7). In addition, the inserted DNA in four of these plasmids (pDm688-132, pDm688-142, pDm688-169, and pDm688-194) was shown capable of pairing with the cloned satellite DNA in pDm688-23 (Carlson & Brutlag, 1977) by heteroduplex analysis (data not shown). The HaeIII resistance of these long cloned satellite DNA segments constitutes evidence that long regions of 1-688 satellite DNA lack HaeIII cleavage sites due to sequence variation.

(b) Homology of HaeIII-resistant 1-688 satellite DNA to satellite DNA containing HaeIII sites

The long HaeIII-resistant satellite DNA fragments cross-hybridize with cloned 1-688 satellite DNA containing HaeIII, HindII and two AluI sites every 359 base-pairs (Carlson & Brutlag, 1977), but the sequences of these two components differ at least to the extent revealed by the frequency of HaeIII sites. As a measure of sequence relationship, the distribution of HindII and AluI sites in HaeIII-resistant DNA was examined. HaeIII-resistant fragments isolated from satellite DNA were digested with HaeIII, HindII and AluI endonucleases. The digestion products were electrophoresed in an agarose gel (Fig. 6(a)), transferred to nitrocellulose by the technique of Southern (1975), and hybridized with radioactively labeled cloned 1-688 satellite DNA. The autoradiogram of the nitrocellulose, shown in Figure 6(b), confirms that the HaeIII-resistant satellite DNA fragments are not cleaved by HaeIII endonuclease. However, digestion with HindII or AluI endonuclease produced a fragment pattern qualitatively similar to that generated from total 1-688 satellite DNA by HaeIII endonuclease digestion. This result indicates that many of the HaeIII-resistant 1-688 satellite DNA fragments contain regularly spaced recognition sites for HindII and AluI endonucleases.

The distribution of HindII sites in individual cloned segments of HaeIII-resistant satellite DNA was also investigated. The gel in Figure 7(a) displays the fragments generated by HindII cleavage of the six largest hybrid plasmids described previously. The fragments were transferred to nitrocellulose (Southern, 1975) and hybridized with a radioactively labeled cloned 1-688 satellite DNA probe. Fragments homologous
Fig. 5. Cloned HaeIII-resistant fragments lack HaeIII endonuclease sites.
Fourteen hybrid plasmids containing HaeIII-resistant 1-888 satellite DNA were digested with HaeIII endonuclease, and the resulting fragments were electrophoresed in 1-4% agarose gels. The fragment patterns of 8 representative hybrid plasmids are shown. As mentioned in the text, the inserted DNA in one hybrid plasmid, shown in slot 3, gives rise to 2 fragments. The fragments common to all the hybrid plasmids are derived from the pSC101 vector. The fragments were transferred from this gel to nitrocellulose (Southern, 1975) and hybridized with 32P-labeled eDm888-52. Autoradiography confirmed that all the non-vector fragments of the 14 plasmids exhibit homology to 1-888 satellite DNA (data not shown). The inserted HaeIII-resistant segments in the 6 largest plasmids (4 of which are shown above) range in size from about 3-5 x 10⁶ base-pairs to 9 x 10⁶ base-pairs as determined from the mobility of the HaeIII fragments containing these inserted segments on a 0-7% agarose gel (data not shown) relative to EcoRI-digested lambda DNA marker fragments (Thomas & Davis, 1975). In addition to an inserted segment, these HaeIII fragments also contain poly(dA)-poly(dT) and a total of 0-2 x 10⁹ base-pairs of adjacent vector DNA (Carlson & Brutlag, 1977).
Fig. 6. HaeIII-resistant 1-688 satellite DNA contains HinfI and AluI endonuclease sites.
(a) Long HaeIII-resistant DNA fragments, purified as described in Materials and Methods, were digested with AluI, HinfI and HaeIII endonucleases. For comparison, total 1-688 satellite DNA was digested with HaeIII endonuclease. The digestion products were electrophoresed in a 1% agarose gel, with equal amounts of DNA loaded on the 3 slots at the left. Although no discrete cleavage products are visible in the AluI and HinfI digests of the HaeIII-resistant fragments, in each case there is clearly less DNA migrating at the limiting mobility than in the HaeIII digest. Since the HaeIII-resistant fragment preparation included contaminant DNA as well as 1-688 satellite DNA, fragments homologous to 1-688 satellite sequences were detected by hybridization, as described in (b).
(b) The fragment pattern was transferred from the gel shown in (a) to nitrocellulose (Southern, 1975). The fragments were then hybridized with 32P-labeled cDm688-62 (9 × 10⁶ cts/min), a probe specific for 1-688 satellite DNA. An autoradiogram of the nitrocellulose is shown. AluI endonuclease digestion may produce fewer distinct bands than does HinfI endonuclease digestion because many 359 base-pair units have 2 AluI sites and only 1 HinfI site.

to the probe were detected by autoradiography (Fig. 7(b)). The fragments from three hybrid plasmids include satellite monomers approximately 359 base-pairs in size. However, many fragments generated from the cloned HaeIII-resistant satellite DNA are much larger than monomer, and also many fragments are not integral multiples of 359 base-pairs.

By comparison of Figure 7(a) and (b), it is apparent that most of the fragments derived from the inserted D. melanogaster DNA segments contain sequences homologous to 1-688 satellite DNA. However, a few fragments, such as the fragment of pDm688-142 mentioned in the legend to Figure 7, did not hybridize with the probe under the hybridization conditions described in Materials and Methods, nor under
Fig. 7. Fragments produced from cloned HaeIII-resistant DNA by HinfI endonuclease digestion.
(a) The following hybrid plasmids containing HaeIII-resistant 1-688 satellite DNA were digested with HinfI endonuclease: pDm688-132, pDm688-194, pDm688-169, pDm688-168, pDm688-142, and pDm688-117. The fragments were separated by electrophoresis in the 1-2% agarose gel shown. Two fragments in each digest contain D. melanogaster DNA joined by poly(dA)·poly(dT) to pSC101 vector DNA; the vector segments consist of the DNA between the EcoRI endonuclease cleavage site of pSC101 and the adjacent HinfI sites. The positions of the 2 vector fragments bounded by EcoRI and HinfI cleavage sites are indicated. The positions of 1-688 satellite monomer and dimer fragments, included on this gel as markers, are also indicated.
(b) The pattern of DNA fragments was transferred from the gel shown in (a) to nitrocellulose (Southern, 1975) and then 32P-labeled cDm688-52 DNA (5 x 10^6 cts/min) was hybridized to the DNA fragments. Fragments homologous to 1-688 satellite DNA were detected by autoradiography, as shown above. The fragments labeled less heavily in proportion to their size are those which contain, in addition to satellite DNA, poly(dA)·poly(dT) and adjacent vector DNA. Most of the fragments derived from the inserted HaeIII-resistant DNA segment exhibit some homology to 1-688 satellite DNA. As mentioned in the text, a few fragments do not hybridize to satellite DNA probe, for example, the second largest fragment of pDm688-142.
less stringent conditions (\(t_m = 40^\circ\text{C}\); data not shown). Such fragments contain D. melanogaster DNA adjacent to a block of 1-688 satellite DNA in the chromosomes (Carlson & Brutlag, 1978).

The same six hybrid plasmids were also digested with AluI endonuclease. Digestion of total 1-688 satellite DNA with this enzyme produces mainly fragments 174, 185, and 359 base-pairs in size (Fig. 8; Hsieh & Brutlag, 1979). Five of the plasmids gave rise to fragments homologous to 1-688 satellite DNA but slightly smaller than these sizes, as judged by their mobility on a 2\% agarose gel; moreover, many large fragments homologous to 1-688 satellite DNA were also generated, as was the case for HinfI digestion of these plasmids (Fig. 7). Unexpectedly, AluI digestion of the sixth hybrid plasmid, pDm688-169, revealed a repeating unit of 254 base-pairs in the inserted satellite DNA.

(c) A region of 1-688 satellite DNA with a 254 base-pair repeating unit

The HaeIII-resistant satellite DNA segment of pDm688-169 is cleaved almost entirely into 254 base-pair fragments by AluI endonuclease (Fig. 8). The inserted segment is \(8.4 \times 10^3\) base-pairs in size (see the legend to Fig. 8) and must therefore contain about 33 tandem 254 base-pair repeats. Some degree of sequence homology between these 254 base-pair repeating units and cloned 359 base-pair units was indicated by both the hybridization observed between pDm688-169 and cDm688-52 (Fig. 8(b)) and the formation of heteroduplex structures by pDm688-169 and pDm688-23, mentioned above (data not shown). Sequence analysis of the 254 base-pair repeat was undertaken to determine the extent of homology with the known sequence of the 359 base-pair unit (Hsieh & Brutlag, 1979).

To prepare radioactively labeled fragments suitable for sequence analysis, we took advantage of several HinfI sites present in pDm688-169; HinfI endonuclease cleaves the inserted satellite DNA at three sites (see Fig. 7(a)), and one of the resulting fragments is 254 base-pairs in size. For analysis of the nucleotide sequences adjacent to AluI sites, the 254 base-pair fragments produced by AluI endonuclease cleavage were labeled with \(^{32}\text{P}\) at the 5' ends. Subsequent cleavage with HinfI endonuclease allowed separation of the two labeled ends of those 254 base-pair fragments containing a HinfI site. For analysis of the sequences flanking the HinfI sites, the single 254 base-pair fragment produced by HinfI endonuclease digestion was end-labeled and then cleaved with AluI endonuclease to separate the two labeled ends. Nucleotide sequences were determined by the procedure of Maxam & Gilbert (1977) (Fig. 9).

Figure 10 compares the sequence of this 254 base-pair repeat with that of a 359 base-pair repeat unit from aDm688-23-24 (Hsieh & Brutlag, 1979). The two repeat units are remarkably similar except for a sequence of 98 base-pairs present only in the 359 base-pair unit. Apart from this sequence, the two monomers are about 80\% homologous and differ mainly by single base insertions, deletions, or changes. The homology between these units suggests that these two types of 1-688 satellite DNA with different repeat sizes evolved from a common ancestor. The isolation of separate tandem repeats of the 254 base-pair unit (33 tandem repeats) and the 359 base-pair unit (16 tandem repeats) in hybrid plasmids pDm688-169 and pDm688-23 (Carlson & Brutlag, 1977) indicates that these two different repeats compose separate regions of 1-688 satellite DNA, and suggests that alteration of the ancestral sequence was followed by amplification of distinctive repeat units, resulting in new homogeneous tandem arrays.
Fig. 8. Fragments produced from cloned *Hpa*III-resistant DNA by *Alu*I endonuclease digestion.

(a) The following plasmids were digested with *Alu*I endonuclease: pSC101, pDm688-117, pDm688-142, pDm688-168, pDm688-194, pDm688-132, and pDm688-169. Purified 1·688 satellite DNA was also digested to provide markers. The digestion products were electrophoresed in a 2% agarose gel. Because the satellite monomer units from the hybrid plasmids migrated faster than the 359 base-pair monomers from purified satellite DNA, a control sample containing a mixture of pSC101 and 1·688 satellite DNA fragments was co-electrophoresed. This control confirmed that the difference in migration was not an artifact of the gel (data not shown). bp, base-pairs.

Continued on page 497
4. Discussion

In these studies we have used molecular cloning to show that nucleotide sequence, rather than base modification, is responsible for the absence of HaeIII endonuclease sites in the dimer fragments and long HaeIII-resistant fragments generated from 1-688 satellite DNA by HaeIII endonuclease digestion. This finding indicates that regions of 1-688 satellite DNA with different distributions of restriction endonuclease recognition sites (Carlson & Brutlag, 1977) differ in nucleotide sequence rather than degree of base modification. In contrast to our results, Gautier et al. (1977) have demonstrated that base modification masks restriction sites in cow satellite DNA. However, the sequence CpG, which is preferentially modified in cow and other DNAs (Sinsheimer, 1955), is not present in any of the restriction sites in 1-688 satellite DNA.

We have also characterized the long HaeIII-resistant regions of satellite DNA to determine their sequence relationship to regions of 1-688 satellite DNA containing HaeIII sites in nearly every repeat. We had previously shown that the HaeIII-resistant portion of the 1-688 satellite would cross-hybridize with cloned monomer repeats (Carlson & Brutlag, 1977). Here we have shown that much of this HaeIII-resistant DNA still contains HindIII and AluI sites distributed at approximately 359 base-pair intervals, or integral multiples thereof. Thus, as judged from the distribution of restriction endonuclease sites, most of the HaeIII-resistant satellite DNA appears to differ from the bulk of the 1-688 satellite DNA mainly in its lack of HaeIII endonuclease sites. However, comparison of the five cloned HaeIII-resistant DNA segments containing repeat units of about 359 base-pairs with the cloned satellite DNA segment in hybrid plasmid pDM688-23 (Carlson & Brutlag, 1977) shows that the sequence variation between different stretches of satellite DNA is in some cases considerable. In contrast to the lack of HaeIII sites and somewhat irregular distribution of HindIII and AluI endonuclease sites in the five cloned HaeIII-resistant segments, pDM688-23 contains a HaeIII, HindIII and two AluI endonuclease sites in each of 16 tandem 359 base-pair units.

An example of variation not only in distribution of restriction sites but also in size of the repeat unit is provided by the HaeIII-resistant segment cloned in pDM688-169. This plasmid contains 33 tandem repeats of a unit only 254 base-pairs in size. Sequence analysis showed that this 254 base-pair unit is homologous to all but 98 base-pairs of the 359 base-pair repeat. Although pDM688-169 offers a particularly striking example of variation in size of the repeat unit, the finding that the other five cloned segments have repeat units slightly smaller than 359 base-pairs suggests that this kind of sequence variation may be common in 1-688 satellite DNA.

Although one of six HaeIII-resistant satellite clones analyzed contained 254 base-pair units, regions composed of 254 base-pair repeat units constitute only a small fraction of the 1-688 satellite DNA. When total 1-688 satellite DNA was cleaved with

(b) The DNA fragments were transferred from the gel to nitrocellulose and hybridized with ^32P-labeled CDm688-52 DNA (1 × 10^6 cts/min). An autoradiogram of the nitrocellulose is shown. Transfer of very small fragments to nitrocellulose is variable in efficiency, and in other experiments we have detected in each hybrid plasmid except pDM688-169 a pair of labeled fragments, corresponding to the pair visible on the gel, of sizes close to 180 base-pairs. Plasmid pDM688-169 generates a strong band of 254 base-pair fragments homologous to 1-688 satellite DNA.

The size of the inserted segment in pDM688-169 was determined to be 8-4 × 10^5 base-pairs by co-electrophoresis of HaeII-digested pDM688-169 DNA with marker fragments in a 0.7% agarose gel (data not shown). HaeII endonuclease digestion of pDM688-169 generates a fragment containing the inserted DNA segment flanked by poly(dA)·poly(dT) and a total of 0-3 × 10^5 base-pairs of adjacent vector DNA (Carlson & Brutlag, 1977).
Fig. 9. Representative autoradiogram of a sequencing gel.
A 168 base-pair fragment end-labeled at the *Hinf*I endonuclease site and including nucleotides 199 to 254 and 1 to 112 was prepared as described in Materials and Methods. The chemical reactions for guanine, strong adenine/weak guanine, cytosine, and thymine and cytosine cleavage were carried out (Maxam & Gilbert, 1977). The reaction products were electrophoresed in both a 20% polyacrylamide gel (a) and a 12% polyacrylamide gel (b) as described in Materials and Methods. Autoradiographs of the gels are shown, and numbers indicate the position in the sequence, as written in Fig. 10, of the nucleotide to which each band corresponds. The sequence read from these gels is complementary to the sequence shown in Figure 10.
Fig. 10. Comparison of the 254 and 359 base-pair repeats of I-688 satellite DNA.

The 254 base-pair sequence shown is derived from those 254 base-pair units in pDm688-169 which contain HinfI endonuclease sites, and the 359 base-pair sequence was determined by Haeh & Brutlag (1979) for a single cloned monomer of satellite DNA. The symbol X in the sequence indicates that the nucleotide has not been determined, and underlining indicates that the sequence is uncertain. The sites of HinfI and AluI endonuclease cleavage of the 254 base-pair repeat are shown. Three instances of sequence heterogeneity among the 254 base-pair units were detected as discrepancies between the sequence obtained from fragments end-labeled at AluI endonuclease sites and that from fragments end-labeled at HinfI endonuclease sites. The alternate nucleotide is indicated at positions 158, 162, and 179. Other sequence variations may have gone undetected. Hyphens have been omitted for clarity. bp, base-pairs.
*Alu* I and fractionated by polyacrylamide gel electrophoresis, DNA fragments 254 base-pairs in size were a very minor product in comparison with those derived from the 359 base-pair repeat (data not shown).

Like the satellite DNA segment in pDM688-23, the $8.4 \times 10^8$ base-pair satellite DNA segment in pDM688-169 appears relatively homogeneous by restriction site mapping; it contains an *Alu* I site in every repeat, and the major heterogeneity detected is the presence of three *HinII* sites within the segment. Moreover, the only heterogeneity detected among those 254 base-pair units used for sequence analysis were single base changes. Comparison of the regions of satellite DNA cloned in pDM688-23 (Hsieh & Brutlag, 1979) and pDM688-169 shows that the sequence variations within each region consist primarily of occasional base changes between neighboring repeats, while the variations between units of the two different regions are both more numerous and more major. These data, together with previous work (Carlson & Brutlag, 1977), imply regional homogeneity in I-688 satellite DNA.

The similarity between the 254 and 359 base-pair units suggests that they evolved from a common ancestral repeating sequence. The existence of tandem arrays of each of these units would then imply that amplification followed their divergence with respect to size. Regional homogeneity may be a direct result of the amplification mechanism.

We thank Michael Goldberg for advice on DNA sequencing techniques. This work was supported by a Basil O'Connor Starter Grant from the National Foundation March of Dimes and by a grant from the National Institute of General Medical Sciences.

One of us (M. C.) was a United States Public Health Service Trainee.

REFERENCES


