Chapter 4

THE HIGHLY REPEATED DNA SEQUENCES OF *DROSOPHILA MELANOGASTER*

DOUGLAS L. BRUTLAG

Stanford University Medical Center, Stanford, California 94305, U.S.A.

The highly repeated DNA sequences of eukaryotes can usually be isolated as satellites in CsCl buoyant gradients. The tandem arrangement of short repeated sequences gives rise to the distinctive physical properties of satellites. These DNAs are usually restricted to centromeric heterochromatin [1,2] and, as might be expected, are neither transcribed nor translated. The function of centric heterochromatin is difficult to assess because, like satellite DNA itself, it is so variable between closely related species. Centric heterochromatin in *Drosophila* was thought to be genetically inert due to the lack of genes in this region and the viability of large heterochromatic deletions [3]. However, such deletions in the sex chromosomes do result in improper disjunction in the first meiotic division and cause a marked decrease in fertility [4]. This genetic evidence indicates that centric heterochromatin and perhaps satellite DNAs are involved in normal chromosome propagation in meiosis. We have, therefore, been studying the arrangement of satellite DNA in the heterochromatin in *Drosophila* in order to provide a molecular basis for such a proposed function.

We have found that the bulk of the heterochromatic DNA of *Drosophila* consists of four different satellite species [5]. These DNAs contain short nucleotide sequences tandemly repeated in arrays over 1,000,000 base pairs long [6]. Several of these satellites, although appearing homogeneous by many physical criteria, contain more than one distinct DNA with different repeating sequences [7]. The sequences in one satellite are so similar that the different DNAs cannot be separated by classical procedures. I will conclude this paper by summarizing the progress we have made in separating these components by cloning individual molecules in hybrid bacterial plasmids.

Drosophila is an ideal organism with which to work since most of the heterochromatic DNA can be isolated in the form of satellite DNA. Fig. 1 shows nuclear DNA centrifuged to equilibrium in a CsCl gradient containing the antibiotic actinomycin D. Three classes of highly repeated DNA do not bind

The Molecular Biology of the Mammalian Genetic Apparatus:
edited by P. Ts'o

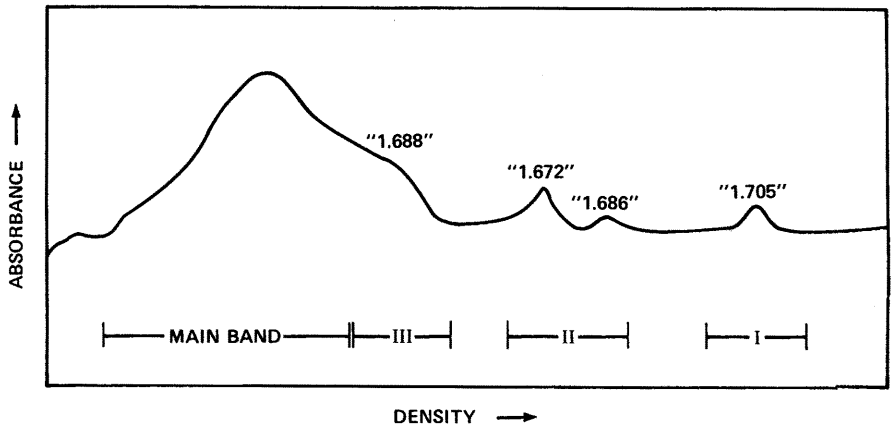

Fig. 1. Analytical actinomycin D-CsCl gradient of *Drosophila melanogaster* DNA. This DNA was isolated from diploid nuclei of developing embryos (9 h mean age).

actinomycin D, and their densities are not altered by the drug. The 1.705, 1.686 and 1.672 species are, therefore, well separated from the main band which is shifted from 1.701 g/cm^3 to 1.62 g/cm^3. A fourth satellite 1.688, which normally co-bands with the 1.686 g/cm^3 species does bind actinomycin D, allowing the resolution of these two satellites. The 1.688 g/cm^3 material can be purified to near homogeneity by repeated centrifugation in CsCl gradients.

Repurification of each species results in DNA which appears homogeneous in both neutral and alkaline CsCl gradients. In alkalinc CsCl each satellite separates into its complementary strands (fig. 2). Note the extreme separation of heavy and light strands of the 1.672 and 1.686 species which has facilitated sequence studies.

These four classes of DNA constitute 18% of the *Drosophila* genome and are located exclusively in centromeric heterochromatin as determined by in situ hybridization [5]. Three of these satellites have been isolated in pure form as molecules 225,000 base pairs in length [6]. From the yield of this isolation, and assuming that satellite regions have a uniform size, we can calculate the minimum length of such a region to be 1,000,000 base pairs. Since the heterochromatin of the X chromosome contains perhaps 8,000,000 base pairs, a block of 1,000,000 satellite base pairs could constitute a major cytological subdivision.

We have carried out partial nucleotide sequence analysis of these species to determine the basic repeated segment (fig. 3). The lightest satellite contains two different repeating sequences [7]. Evidence for this is three-fold. First, pancreatic RNase digests of RNA complementary to the T rich strand give AAU and AU as major products. Partial ribonuclease digests show the major repeating units are these five and seven nucleotide sequences. Longer oligonucleotides with two and three of these five nucleotide repeats in a row have been found. Similarly,

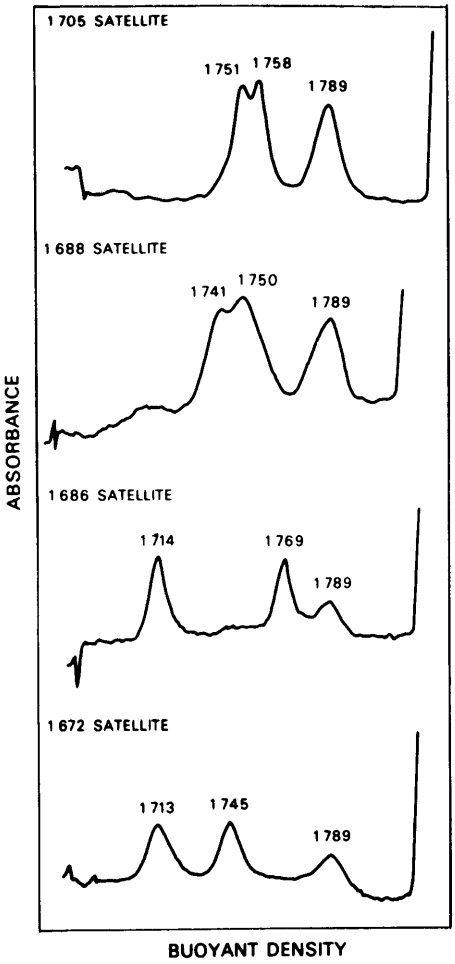

Fig. 2. Alkaline CsCl gradients of purified satellite DNA fractions. *M. luteus* DNA was added as a density marker in each gradient (ρ = 1.789 g/cm^3).

oligonucleotides containing two and three repeats of the seven nucleotide sequence have been found.

A second line of evidence for two distinct DNAs comes from thermal denaturation studies on the separated complementary strands. Since each sequence is almost an alternating AT, we predicted that both the heavy and light strands might undergo base pairing with itself in the form of a hairpin. One sequence could pair with itself but with one TT mismatch every five nucleotides. Likewise, the other sequence would pair with itself with a TT mismatch every seven nucleotides. Electron microscopy of the separated strands showed that they were double helical and thermal denaturation of each showed a biphasic melt. These transitions occured 15° and 23° below the T_m for a perfectly paired

SEQUENCES OF HIGHLY REPEATED DNA

SATELLITE	SEQUENCE	COMPOSITION
1.672	A A T A T T T A T A	60%
	A A T A T A T T T A T A T A	40%
1.686	A A T A A C A T A G T T A T T G T A T C	80%
	$\begin{pmatrix} A_5\ T_3\ C_1\ G_1 \\ T_5\ A_3\ G_1\ C_1 \end{pmatrix}$	10%
	$\begin{pmatrix} A_7\ T_1\ C_1\ G_1 \\ T_7\ A_1\ G_1\ C_1 \end{pmatrix}$	10%
1.688	Hae III SITE $\begin{pmatrix} G\ G\ C\ C \\ C\ C\ G\ G \end{pmatrix}$ DISTRIBUTED REGULARLY, OFTEN 365 NUCLEOTIDES APART	
1.705	A A G A G T T C T C	40%
	A A G A A G A G A G T T C T T C T C T C	~ 25%

Fig. 3. Sequences detected in satellite DNA [5,7,9,10].

d(AT), consistent with two DNA species, one with a mismatch every five base pairs and one with a mismatch every seven.

A third experiment indicates that these sequences are on different molecules. If the A rich strand of this satellite is used as a template for DNA polymerase, pyrimidine tract analysis of the product showed both sequences were copied faithfully and confirmed the RNA sequencing results. If, however, dGTP and dCTP were left out of the DNA polymerase reaction, the amount of product was reduced 50% and pyrimidine tract analysis showed only the five nucleotide repeat was synthesized. We conclude that molecules containing the five nucleotide

repeat are highly uniform, but the seven nucleotide repeat may have a rare G or C residue distributed throughout. Base composition indicates G and C are present at the 0.5% level.

Sequence analysis of the 1.686 satellite shows three repeating sequences. When RNA complementary to the heavy strand of this satellite is cleaved with T_1 ribonuclease, three oligonucleotides ten long are produced. The major component has the sequence 5′ AAUAACAUAG 3′ determined by sequential digestion with spleen phosphodiesterase. The two minor sequences have differing base compositions. We do not know if these different sequences are tandemly arranged or if they are present as distinct DNA species.

The 1.688 DNA is the most difficult to purify and sequence analysis has only begun. Manteuil et al. [8] have shown that a particular restriction site (the *Hae*III sequence 5′GGCC 3′) is distributed in a regular fashion in this satellite. Work in our laboratory confirms this and preliminary sequence analysis indicates the basic repeating sequence of this DNA is much more complex than the others.

The sequence of the 1.705 satellite has been reported in three labs. The sequence AAGAG has been reported by Sederoff et al. [9] and by Endow et al. [10]. These workers used only ATP and GTP to transcribe complementary RNA. We have used all four triphosphates and have detected a second repeating sequence in the RNA produced. These sequences were analyzed by partial T_1 ribonuclease followed by sequential digestion of the partial products.

It is clear that separation of such closely related DNAs would be difficult, if not impossible, by classical physical techniques. Two repeating sequences such as those present in 1.705 g/cm^3 species have identical base compositions, nearest neighbors and nucleotide strand bias. Moreover, for the study of the arrangement of these sequences in the centric heterochromatin, it is essential to have all sequences separated and in pure form. We have, therefore, made synthetic hybrid bacteria plasmids [11], each containing a single molecule of satellite DNA. These hybrids will allow us to separate each buoyant class into its various molecular components in order to simplify sequencing and to provide hybridization probes to locate each sequence in situ.

Insertion of satellite DNA into bacterial plasmids is not straightforward. Since none of the *Drosophila* satellites contain sites for R1 endonuclease, one must synthesize complementary terminal sequences on the plasmid and the satellite. This is normally accomplished by shearing the DNA and then treating it with a 5′ exonuclease to expose single-stranded 3′ termini to act as primers for terminal transferase [12]. Treating satellite DNA with a 5′ exonuclease exposes complementary 3′ sequences, and the molecules immediately cyclize and the ends become unavailable for terminal addition. We have, therefore, found conditions for terminal addition to double-stranded DNA. Basically, by lowering the ionic strength of the terminal transferase reaction, we have favored fraying at the end

of a helix which permits the enzyme to add dA tails to double-stranded satellite DNA. Similar conditions allow addition of dT tails directly to R1 cleaved plasmid DNA without 5′ exonuclease treatment (fig. 4).

We have then annealed the satellite and plasmid DNA in the normal way and transfected bacteria deficient for recombination, restriction and modification (fig. 4). Colonies resistant to tetracycline, a gene carried on the plasmid DNA, were selected. Among these we have further selected colonies containing satellite DNA sequences by hybridization of RNA complementary to the satellite (1.705 in this case) to the plasmid DNA of the bacterial cells [13].

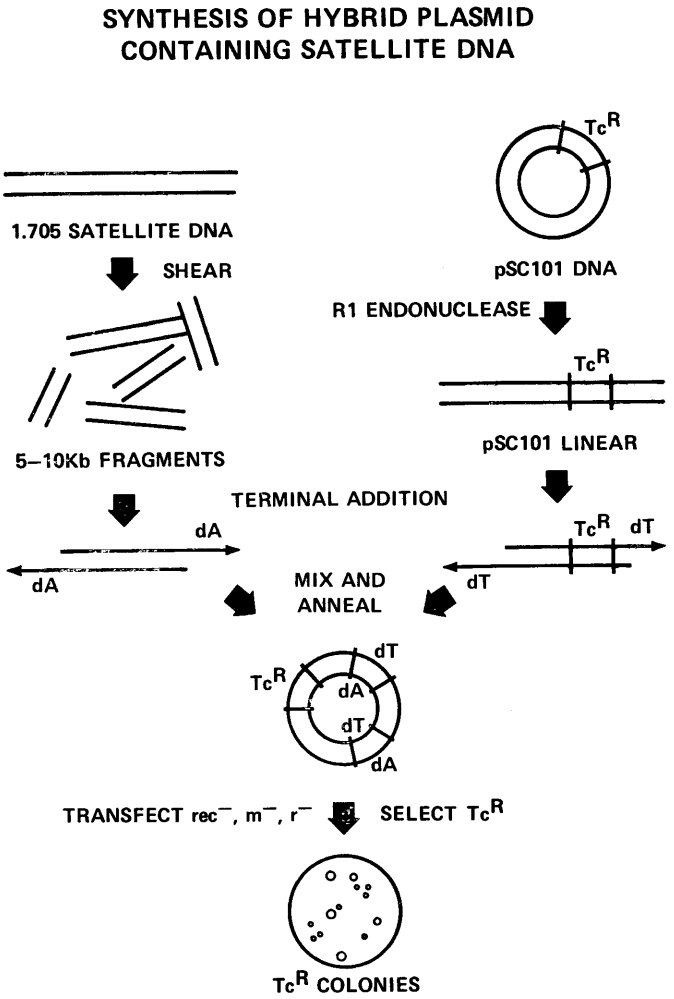

Fig. 4. Scheme for synthesis of hybrid bacterial plasmids containing satellite DNA. Basically this procedure is as described in [11] with the exception of the 5′ exonuclease step, which was omitted as described in the text.

Only 30% of all the colonies tested contained 1.705 DNA sequences by this hybridization criteria. This was unexpected since the presence of satellite DNA with A tails was absolutely required to obtain Tc^R colonies. The pSC DNA by itself was not infectious. The second unexpected result was the length of the inserted sequence. DNA of 5–10 kb average length was used and was present in annealed molecules, but the average insert in the plasmids was usually less than 1 kb. Occasionally, individual clones contained plasmid DNA heterogeneous in size. Upon subcloning, stable and homogeneous plasmids are obtained but with small 1 Kb satellite inserts. Based on the instability of large satellite inserts and the stability of shorter satellite regions, we have concluded that intramolecular recombination is occurring within tandemly repeated regions. We have used both $recA^-$ and $recB^-C^-$ double mutants. This instability may be due to small amounts of residual recombination in these strains, combined with the traditional instability of tandemly repeated sequences in *E. coli*, or it may be due to a strong selection against the presence of such unusual sequences.

With plasmids which have probably undergone deletion and rearrangement, it is clear that one cannot study long range sequence regularities such as the organization of restriction sites in satellite DNA. However, since each plasmid has derived from a single molecule, we are now in a position to isolate various molecular components of each buoyant class by hybridization. We will also be able to determine the basic repeating units contained within each molecule and to detect other closely related sequences in the satellites which may have escaped our sequence analysis. By using complementary RNA prepared from these plasmids as probes for in situ hybridization, we will test the hypothesis of a chromosome specific location of each of the different repeated sequences [7].

Acknowledgements

This work was supported in part by the National Institutes of Health (GM 21498-01) and the National Foundation March of Dimes.

The sequence work reported in this paper was performed in collaboration with Dr. W.J. Peacock in his laboratory in Canberra, Australia. The work on the plasmids was in association with Dr. Kirk Fry, Timothy Nelson, Marian Carlson and Peggy Hung.

References

[1] M.L. Pardue and J.G. Gall, Chromosomal localization of mouse satellite DNA. Science 168, 1356 (1970).

[2] K.W. Jones, Chromosomal and nuclear location of mouse satellite DNA in individual cells. Nature 225, 912 (1970).

[3] S. Gershenson, Studies on the genetically inert region of the X-chromosome of *Drosophila*. J. Genetics 28, 297 (1933).
[4] W.J. Peacock and G.L.G. Miklos, Meiotic drive in *Drosophila*: New interpretations of segregation distorter and sex chromosome systems. Adv. Genetics 17, 361 (1973).
[5] W.J. Peacock, D. Brutlag, E. Goldring, R. Appels, C.W. Hinton and D.L. Lindsley, The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. Cold Spring Harbor Symp. Quant. Biol. 38, 405 (1973).
[6] E.S. Goldring, D.L. Brutlag and W.J. Peacock, Arrangement of highly repeated DNA of *Drosophila melanogaster*, in: The Eukaryote Chromosome (W.J. Peacock and R.D. Brock, eds.), Australian National University Press, Canberra, Australia, 1975, p. 47.
[7] D.L. Brutlag and W.J. Peacock, Sequences of highly repeated DNA in *Drosophila melanogaster*, in: The Eukaryote Chromosome (W.J. Peacock and R.D. Brock, eds.), Australian National University Press, Canberra, Australia, 1975, p. 35.
[8] S. Manteuil, D.H. Hamer and C.A. Thomas, Jr., Regular arrangement of restriction sites in *Drosophila* DNA. Cell 5, 413 (1975).
[9] R. Sederoff, L. Lowenstein and H.C. Birnboim, Polypyrimidine segments in *Drosophila melanogaster* DNA: II Chromosome location and nucleotide sequence. Cell 5, 183 (1975).
[10] S.A. Endow, M.L. Polan and J.G. Gall, Satellite DNA sequences of *Drosophila melanogaster*. J. Mol. Biol. 96, 665 (1975).
[11] P.C. Wensink, D.J. Finnegan, J.E. Donelson and D.S. Hogness, A system for mapping DNA sequences in the chromosome of *Drosophila melanogaster*. Cell 3, 315 (1974).
[12] P.E. Lobban and A.D. Kaiser, Enzymatic end to end joining of DNA molecules. J. Mol. Biol. 78, 453 (1973).
[13] M. Grunstein and D.S. Hogness, Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. Proc. Nat. Acad. Sci. U.S.A. 72, 3961 (1975).