© 1992 Oxford University Press Nucleic Acids Research, Vol. 20, No. 5 1113–1116

# A superfamily of *Drosophila* satellite related (SR) DNA repeats restricted to the X chromosome euchromatin

Susan M.DiBartolomeis*, Kenneth D.Tartof and F.Rob Jackson
Neurobiology Group, Worcester Foundation for Experimental Biology, 222 Maple Avenue, Shrewsbury, MA 01545 and [1]The Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, USA

Received November 6, 1991; Revised and Accepted January 16, 1992 EMBL accession nos X62937–X62939 (incl.)

## ABSTRACT

**The 1.688 g/cm$^3$ class of *Drosophila* satellite DNA is predominantly localized to the centromeric heterochromatin of the X chromosome. We report here the existence of 1.688 satellite related (SR) DNA arrays present at numerous locations throughout the euchromatic portion of the X. Unlike their heterochromatic counterparts, euchromatic SRs consist of a small number of repeating units (usually 2–4), each of which is 63–81% identical to the 359-bp monomer of the 1.688 satellite. Although it appears that SR DNA arrays are not transcribed, in at least two cases, they are located adjacent to transcriptionally active genes. SR sequences also have significant similarity to a previously described *Drosophila* middle repeat found almost exclusively in the X euchromatin. It seems likely that these X linked sequences are required for sex chromosome specific functions.**

## INTRODUCTION

A large number of studies have documented the structure and organization of a class of reiterated sequences known as satellite DNA (see references 1 and 2 for recent reviews). Satellites are composed of tandem repeats whose unit structure has been described as simple (2–10 basepairs, bp) or complex (100–700 bp). While the functions of these repeated DNAs have not been rigorously defined, it has been hypothesized that some may confer a particular higher order chromatin structure or be required for chromosome pairing and meiotic recombination.

Four distinct satellite DNAs are known to exist in the *Drosophila melanogaster* genome (1). Among these is a complex class known as the 1.688 g/cm$^3$ satellite, so named for its buoyant density, that comprises about 4% of the *Drosophila* genome. It consists principally of tandem repeats of a 359 bp sequence that is reiterated on the order of $10^4$ times (3). Although there is some length variation of the repeating monomeric unit (there are 353 bp and 254 bp variants that possess internal deletions of the canonical 359 bp structure), most of the sequence differences within arrays of this satellite are single base substitutions (3,4).

The 1.688 g/cm$^3$ class of satellite DNA is predominantly localized to the centromeric heterochromatin of the *Drosophila* X chromosome. During unrelated studies of two different X-linked loci, however, we have discovered that 1.688 satellite related (SR) DNA sequences are present in the euchromatic portion of the genome as well. In previous studies of *white* gene variegation, euchromatic DNA segments that hybridize with 1.688 satellite sequences were identified in a region distal to the *white* gene (5). More recently, two SR DNA arrays were discovered near the *miniature-dusky* (*m-dy*) locus in region 10E1-2. Similar to the centric 1.688 arrays, the repeating monomer within the euchromatic DNA segments is approximately 359 bp. However, unlike the heterochromatic satellite, SRs consist of short tandem clusters of only 2 to 4 repeating units. While they are not themselves transcribed, in at at least two cases, they are located immediately adjacent to transcriptionally active genes (6). We also note that the SR sequences described here are related to a previously reported *Drosophila* middle repetitive element which is also dispersed along the X (7).

## EXPERIMENTAL PROCEDURES

### Isolation of Phage Clones

Overlapping phage clones containing genomic inserts from the 10E1-2 region were isolated from a l Charon 4A/*Drosophila* library (8) and mapped with restriction endonucleases (9). The l phage clones from the 3C region were previously described (5). One of these, clone M505, encompasses a 1.5 kb segment that hybridizes to the 1.688 heterochromatin satellite sequence.

### Chromosome *In Situ* Hybridization

*In situ* hybridizations to salivary gland polytene chromosomes were performed according to published procedures (9) to confirm the cytogenetic location of the isolated phage clones as well as to identify the multiple sites to which a 3C SAR sequence hybridized.

* Present address: Department of Biology, Millersville University, Millersville, PA 17551, USA

### DNA Sequence Analysis

DNA sequences were determined using standard methods (10) with either [$^{32}$P]dCTP and the Klenow fragment of DNA polymerase or [$^{35}$S]dATP (11) and Sequenase (USB). Genomic fragments were subcloned into the pBluescript plasmid vector or M13 mp9 for DNA sequencing studies. Sequence comparisons were performed using the programs of The Genetics Computer Group (12) implemented on a VAX 11/750 with a VMS 5.3 operating system.

## RESULTS

### Organization of the Euchromatic SR DNA Segments

Figure 1 depicts the organization of the SR DNAs present in regions 3C and 10E. Within region 3C, at approximately −24 kb on the molecular map of the *white* locus, there exists a 1.4 kb genomic segment that has previously been shown to hybridize to 1.688 satellite sequences (5). We refer to this particular 1.688 SR segment as 1688-3C to denote both its homology to satellite sequences and its location at cytogenetic position 3C on the polytene chromosome map (Figure 1A). This genomic segment contains a tandem array of three sequence blocks, each of which has significant similarity to the 359 bp repeating unit present in 1.688 satellite. Two of the 3C repeats are full length, whereas the third is shorter than a satellite monomer. The entire 1688-3C cluster of repeats is flanked at each end by a direct repeat of the tetranucleotide, GCAT. Target site duplications of this sort are frequently associated with transposon insertion, and their presence here suggests this as the means by which SR DNA sequences may be mobilized.

A portion of the *m-dy* region within 10E is shown in Figure 1B. Two divergently oriented SR arrays, 1688-10E$^{d}$ (distal) and 1688-10E$^{p}$ (proximal), are present and separated by about 24 kb of intervening DNA. The distal cluster is composed of two SR monomeric units, one complete and one incomplete, while the proximal array contains three complete units and one partial repeat. As in region 3C, the basic repeating unit is slightly longer than 350 bp, similar to the repeating unit of centromeric satellite. The distal or proximal 10E SR arrays also appear to be flanked by tetranucleotide (AATT) or trinucleotide (AAT) direct sequence repeats, respectively. A curious feature of all of the euchromatic SR sequences is a periodicity of *Eco*RI restriction sites within the repeating units. Five of the six repeats at 10E and two of the three copies at 3C contain *Eco*RI sites.

Finally, it should be noted that none of the SR arrays appears to be transcribed. However, RNAase protection and Northern blot analyses indicate that both 1688-10E$^{d}$ and 1688-10E$^{p}$ are situated within 1 kb of transcribed genomic intervals (6 and data not shown).

### Sequence Relatedness of SR DNAs

Each of the three SR clusters has been completely sequenced. The monomers within each cluster are 85−98% identical to one another, and 70−80% identical to monomers from other SR clusters or the 1.688 heterochromatic satellite. Searches of the Genbank and EMBL databases with SR euchromatic repeats from the 1688-3C and 1688-10E$^{d}$ clusters yielded several highly significant matches, all of which were members of the 1.688 heterochromatic satellite DNA family. We also discovered that the SR repeats have significant sequence similarity to a *Drosophila* middle repetitive DNA previously described by Waring and Pollack (7). As expected, the Waring-Pollack

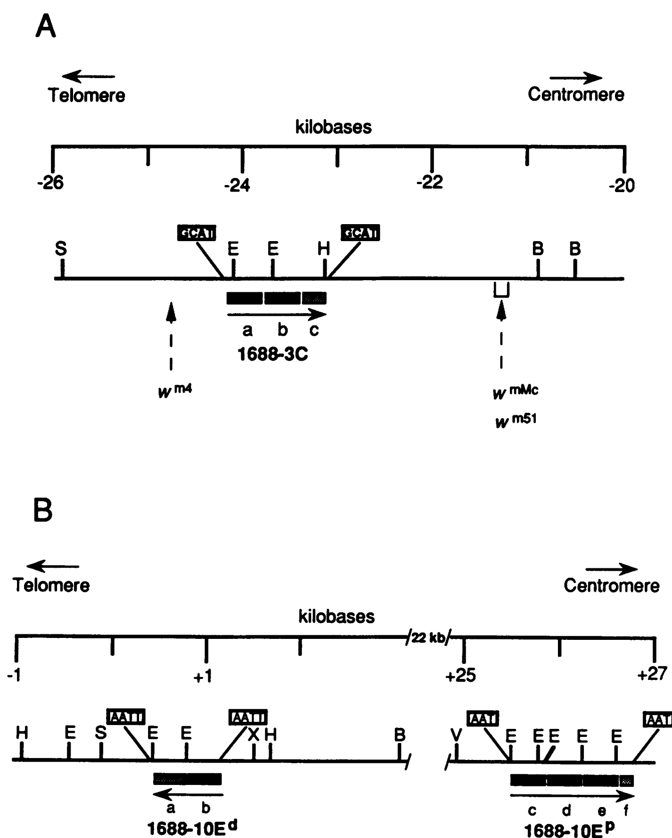


**Figure 1.** Genomic organization of the SR DNA arrays in regions 3C (A) and 10E (B). Molecular positions (in kilobases) correspond to the cloned *white* or *m-dy* regions. The vertical arrows in (A) indicate the positions of three different *white* (*w*) mutations that map near the 1688-3C array. The horizontal arrows under the restriction maps show the orientations of the different SR domains relative to the published sequence of the 1.688 satellite. Black rectangles signify complete SR monomers, whereas shaded rectangles indicate regions with partial homology to 1.688 satellite sequences. The direct terminal repeats flanking the 1688-3C and 1688-10E clusters are indicated. B=BamHI, E=EcoRI, H=HindIII, S=SalI, V=EcoRV, X=XhoI.

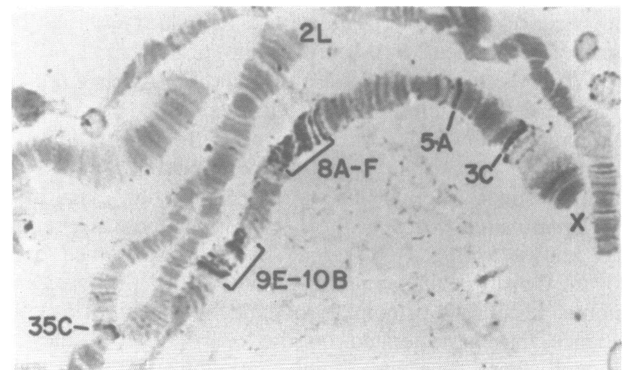


**Figure 2.** *In situ* hybridization of the 1688-3C SR probe to the *Drosophila* X chromosome. Sites of hybridization are indicated by reference to positions on the salivary gland polytene chromosome map. X indicates the X chromosome, and 2L indicates the left arm of chromosome 2. Hybridizations were carried out as previously described (9).

elements also have substantial similarity to the the 1.688 satellite itself, although this has heretofore gone unnoticed. Thus, we now refer to the Waring-Pollack middle repeat in region 11EF as SR 1688-11EF. Bestfit analysis (12) of the 1.688 satellite, and SR monomers 1688-10Ee and 1688-3Cb demonstrate, for example, that they are 68%, 73% and 68% identical, respectively, to 1688-11EF. Furthermore, the 1688-11EF SRs also possess the periodic *Eco*RI sites that are so characteristic of the other SR sequences described above.

### SR Sequences Are Present in Many Regions of the Euchromatin

*In situ* hybridization to salivary gland polytene chromosomes with a probe representing the 1688-3C array demonstrates that similar sequences are present in several different regions of the X chromosome (Figure 2). Conspicuous hybridization is detected at bands in regions 3C, 5A, 8A-F and 9E-10B of the X chromosome and to region 35C of the second chromosome. Importantly, this probe does not hybridize to 10E nor to 11EF, regions in which other SR sequences are known to reside (7, and this paper). That is, the three SR subfamilies at 3C, 10E and 11EF are almost entirely nonoverlapping. It is also noteworthy that the centromere, at which the heterochromatic 1.688 satellite sequences resides, is devoid of hybridization signal when 1688-3C is used as a probe. This is expected since SR sequences in different cytogenetic locations are, at best, only 80% identical to the centric satellite and to one another. The conditions for chromosome hybridization are sufficiently stringent to prevent stable duplex formation when this extent of mismatch occurs. Hence, SR sequences may represent an extended family of repetitive elements related to the 1.688 heterochromatic satellite that are dispersed throughout X chromosome euchromatin.

## DISCUSSION

We have described a family of satellite-related DNA repeats whose distribution is largely restricted to the X chromosome. Although this genomic distribution is intriguing, the function of these sequences remains speculative. With regard to possible functions for SR sequences, it is of interest that in at least one euchromatic region (10E), an SR array is located within 500 bp of a transcribed interval. The sequences might, therefore, bind proteins that are capable of influencing gene regulatory functions. Analogous to true satellite DNAs that are known to be associated *in vivo* with proteins (13,14), perhaps the SR arrays serve as recognition sites for sequence specific DNA binding proteins. Such proteins might, for example, locally influence chromatin structure, thereby inhibiting or activating genes in the vicinity.

Such a regulatory function could be related to the process of X chromosome gene dosage compensation (15). Recent studies (16), for example, have strongly indicated that dosage compensation of at least some X linked genes is effected by changes in local chromosome environment (i.e., chromatin conformation), rather than being brought about by closely linked gene-regulatory sequences. In this regard, it is intriguing that one gene, *mle*, which has been implicated in dosage compensation, encodes a putative DNA helicase that binds to many euchromatic sites along the X chromosome (17).

Alternatively, SR arrays may function in a structural, rather than regulatory, capacity. An interesting prospect in this regard involves the process of X chromosome inactivation that occurs during spermatogenesis in male heterogametic organisms including *Drosophila* (18). This 'precocious X inactivation' appears to be required during spermatogenesis for male fertility. It has been previously noted that the radiation induced euchromatic breakpoints of three *white mottled* mutations flank the 1688-3C array (5). This observation suggested that in the spermatocyte, the X chromosome is folded so that the centric heterochromatin and the *white* region are ectopically paired through their shared satellite related homology (5). It is conceivable this ectopic association actually reflects the ability of SR domains to facilitate a concerted folding of the X, thereby organizing its ultimate inactivation in the spermatocyte.

## ACKNOWLEDGEMENTS

The sequences discussed in this paper have been submitted to the EMBL database under the accession numbers X62937 ($10E^p$), X62938 ($10E^d$), and X62939 (3C). We thank Allan Lohe for discussions and advice concerning *Drosophila* satellite DNA, and Tom Cline and Barry Ganetzky for discussions of dosage compensation. We are grateful to Laurel Newby, Kathy Elliott, Barbara Walker, and Marilyn Bremer for excellent technical assistance during these studies. This research was supported by grants from the NIH (NS30386) and NSF (BNS-8701335) to F.R.J., and by grants to K.D.T. from the NIH (CA-06927, RR-05539), the NSF (9016548) and the Commonwealth of Pennsylvania. S.M.D. was supported by an NIH postdoctoral training grant (T32HD07312).

## REFERENCES

1. Lohe, A. and Roberts, P. (1988). Evolution of satellite DNA sequences in *Drosophila*. In Heterochromatin: molecular and structural aspects. ed., R.S. Verma, Cambridge Univ. Press (Cambridge) pp. 148–186.
2. Willard, H.F. (1989). The genomics of long tandem arrays of satellite DNA in the human genome. Genome 31, 737–744.
3. Hsieh, T.-S. and Brutlag, D. (1979). Sequence and sequence variation within the 1.688 g/cm$^3$ satellite DNA of *Drosophila melanogaster*. J.Mol.Biol. 135, 465–481.
4. Carlson, M. and Brutlag, D. (1979). Different regions of a complex satellite DNA vary in size and sequence of the repeating unit. J.Mol.Biol. 135, 483–500.
5. Tartof, K.D., Hobbs, C. and Jones, M. (1984). A structural basis for variegating position effects. Cell 37, 869–878.
6. DiBartolomeis, S.M. and Jackson, F.R. The *Drosophila dusky* mRNA encodes a putative transmembrane protein that is required at the pupal stage of development for wing epidermal cell differentiation. Submitted for publication.
7. Waring, G.L. and Pollack, J.C. (1987). Cloning and characterization of a dispersed, multicopy, X chromosome sequence in *Drosophila melanogaster*. Proc.Natl.Acad.Sci.USA 84, 2843–2847.
8. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell 15, 687–701.
9. Engels, W.R., Preston, C.R., Thompson, P. and Eggleston, W.B. (1986). *In situ* hybridization to *Drosophila* salivary chromosomes with biotinylated DNA probes and alkaline phosphatase. Focus 8, 6–8.
10. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc.Natl.Acad.Sci.USA 74, 5463–5467.
11. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983). Buffer gradient gels and $^{35}S$ label as an aid to rapid DNA sequence determination. Proc.Natl.Acad. Sci.USA 80, 3963–3965.
12. Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.
13. Levinger, L. and Varshavsky, A. (1982). Protein D1 preferentially binds A+T-rich DNA *in vitro* and is a component of *Drosophila melanogaster* nucleosomes containing A+T-rich satellite DNA. Proc.Natl.Acad.Sci.USA 79, 7152–7156.

14. Strauss, F. and Varshavsky, A. (1984). A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell 37, 889–901.
15. Lucchesi, J.C. and Manning, J.E. (1987). Gene dosage compensation in *Drosophila melanogaster*. Adv.Genet. 24, 371–429.
16. Sass, H., and Melseson, M. (1991). Dosage compensation of the *Drosophila pseudoobscura Hsp82* gene and the *Drosophila melanogaster Adh* gene at ectopic sites in *D. melanogaster*. Proc.Natl.Acad.Sci.USA 88, 6795–6799.
17. Kuroda, M.I., Kernan, M.J., Kreber, R., Ganetzky, B., and Baker, B.S. (1991). The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. Cell 66, 935–947.
18. Lifschytz, E. and Lindsley, D.L. (1972). The role of X-chromosome inactivation during spermatogenesis. Proc. Natl. Acad. Sci. USA 69, 182–186.