Interdigitated arrangement of two oligo(A)-terminated DNA sequences in *Drosophila*

Pier Paolo Di Nocera and Igor B. Dawid

Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20205, USA

Received 6 May 1983; Accepted 22 July 1983

**ABSTRACT**

A cluster of repeated sequences composed of three distinguishable units has been isolated from *Drosophila* melanogaster, and characterized. The region, cloned as pDM158, contains a segment that is homologous to the type 1 ribosomal insertions, a member of the F family of transposable sequences, and a newly described repeated sequence that we have named G. F elements are transposable sequences that lack terminal repeats, generate target site duplications at the point of insertion, and contain an oligo(A) stretch at one end. G sequences are structurally similar though non-homologous to F in that they also carry an oligo(A) stretch. The structure of the 158 region of the genome is best explained by assuming three consecutive events. An F element did insert into a ribosomal insertion-like sequence, followed by the introduction of a G sequence into F. Subsequently, a DNA segment comprising a portion of G and F was tandemly triplicated to yield the arrangement observed. The nested interspersion of repeated sequence elements may be a common feature of eukaryotic genomes.

**INTRODUCTION**

F elements have been characterized as a distinct class of repetitive transposable sequences in *Drosophila* melanogaster (1, 2, 3) lacking the long terminal direct or inverted repeats which occur in other transposable families described so far in this animal (4, 5, 6). The most distinctive property of the F element family is the presence of an oligo(A) stretch at one end of each member, reminiscent of certain pseudogenes (7, 8, 9, 10) and other repeated DNA sequences in mammals (11, 12). F elements are located in the chromocenter and at about 25 euchromatic sites (1, 2, 3). Among the DNA molecules containing F elements that have been analyzed one region, cloned as pDM101, contains segments homologous to the type 1 ribosomal insertion in addition to F (1). Ribosomal insertion-like DNA occurs in the chromocenter of *D. melanogaster*, is usually arranged in tandem copies, and is interspersed with other DNA elements (1, 13, 14, 15). Further analysis of our collection of cloned DNA molecules containing insertion-like sequences revealed that they are frequently interspersed with F elements. In one of these clones an F element...
interrupts an insertion-like sequence similarly to 101F, and further is itself interrupted by other repeated DNA sequences. The interdigitating structure of this DNA region and the possible mechanism underlying such an arrangement are the subject of the present report.

RESULTS
The structure of pDmI 158: F and G sequences
The region cloned as pDmI 158 contains a ribosomal insertion-like sequence interrupted by unrelated repetitive DNA (1). Restriction analysis and hybridization by the Southern blotting procedure led to a map for this DNA region as shown in Figure 1. We compare this structure to the map of pDmI 101 where a complete F element interrupts an insertion-like sequence.

Fig. 1  Map of the D. melanogaster DNA fragment cloned in pDmI 158 and comparison with pDmI 101.
Type I insertion-like sequences are indicated as solid bar, F sequences as hatched bars, and G sequences as open bars. The F element in 101 is subdivided into a, b, and c to correspond to the segments of the F element in 158 which are separated by the insertion of G. The relationship between G1 and G2, 3 is also indicated. See text for further details. Restriction sites are shown as follows: B, Bgl II; S, Sal I; Sm, Sma I; E, Eco RI; X, Xho I; C, Cla I; Ba, Bam HI; H, Hind III.
The insertion-like elements in 101 and 158 are interrupted by F elements at the same position (see also below) and the relative orientation of the two sequences is the same in both cloned regions.

The F element in pDmI 158, which will be called 158F hereafter, is split into three regions by the presence of other sequences. We have indicated the three segments of 158F as a, b, and c, but it must be stressed that this distinction is based only on the structure of this clone, while these three segments are contiguous in 101F and all the other F elements analyzed (3). As indicated in Figure 1 158F is interrupted close to its left end by an unrelated sequence (Gl). After about 3.5 kb of this sequence, the F element continues with the segment called Fb. A further interruption of F sets in about 1 kb to the right with the region G2, repeating part of the Gl segment. Only after another repetition of the G and Fb regions does the bulk of the F element, Fc, continue without further interruption.

Detailed restriction enzyme analysis and partial sequencing (see below) showed that the three copies of the Fb region are identical. The G regions are also very similar, with G2 and G3 identical to the extent analyzed but differing from Gl. G2 and G3 miss the left portion of Gl, but carry an additional sequence of about 400 bp at an internal position introducing a Hind III site which is absent in Gl. The simplest interpretation of this composite structure is that a segment of a repeated DNA family (G sequences) was inserted between the a and b regions of a previously uninterrupted F element; subsequently or simultaneously, a segment of about 3.5 kb containing both G and F sequences has been triplicated. Later changes in G would account for the differences between Gl and G2, 3.

**Interspersed arrangement of F and G occurs in the genome**

We wished to examine whether the F/G interspersion is present in the same arrangement in the genome or whether a cloning artifact might have led to the structure of pDmI 158. To this end genomic Southern blots were carried out (Fig. 2). A short G-specific probe, represented by a Bam HI-Bgl II fragment present at the boundary of the G/F segment amplified in pDmI 158, was nick-translated and hybridized to total Drosophila DNA digested with Bam HI, Bam HI + Bgl II, and Bgl II. Known amounts of similarly digested pDmI 158 DNA were loaded in parallel lanes. In each case the genomic blot contained a band of the same size as the cloned DNA blot. The most diagnostic lanes contain the Bam HI digest, yielding a 1.4 kb fragment that spans the G/F boundary (thin line under the map in Figure 2). This fragment is present in the genome at an intensity corresponding to about 3 copies,
Fig. 2  Genomic Southern blot of F/G structure. Oregon R DNA (2 μg per lane) was digested with Bam HI, Bam HI+Bgl II and Bgl II, separated on 1.2% agarose gel, transferred to nitrocellulose and hybridized to a labeled Bam HI-Bgl II fragment from pDmI 158. Different amounts of similarly digested pDmI 158, corresponding to 1, 3 and 5 copies per genome equivalent, were electrophoresed in adjacent lanes. For each restriction digest the left lane contains genomic DNA and the next three lanes contain cloned material. In the map below the Bam HI (Ba) and the Bgl II (B) sites of pDmI 158 are shown. The probe is underlined by a heavy line, the fragments expected from each digest by thin lines.

suggesting that the particular arrangement in pDmI 158 occurs at only one site.

The genomic blots show that G elements occur at additional locations, not represented by pDmI 158. The total number of G elements has not been determined with accuracy but appears to be between 10 and 20 in the genome of the Oregon R strain used in these experiments.

**Boundary Analysis**

Several regions of pDmI 158 were analyzed by DNA sequencing in order to define precisely the limits and the nature of several boundaries between the different repeated DNA segments. The relevant portions of the nucleotide sequences obtained are presented in Figure 3. Several points of interest result from this analysis. First, 101F and 158F elements interrupt ribo-
Fig. 3 DNA sequences at segment boundaries in pDMI 158 as compared with pDMI 101. The clone from which each sequence is derived is indicated at the left. At the right the boundary is identified by reference to the map below. For a more detailed map see Figure 1. Homologies are indicated by short lines. Dots indicate that the sequence continues but has not been determined in this particular region. The top lines show that 158 and 101 are identical at the insertion-like/F boundary. The target site known to be duplicated by F element insertion in 101 is shown boxed. The next three lines show that 30 bp are missing from the 101F sequence between 158Fa and 158Fb. Further, the sequence to the left of boundaries 4 and 6 shows the oligo(A) stretch and polyadenylation signal (underlined) at the right end of G. The bottom three lines identify boundary 3 by comparison with 5. Two short direct repeats are overlined. Sequencing was done by the method of Maxam and Gilbert (22).

Nucleic Acids Research

Fig. 3 DNA sequences at segment boundaries in pDMI 158 as compared with pDMI 101.

The clone from which each sequence is derived is indicated at the left. At the right the boundary is identified by reference to the map below. For a more detailed map see Figure 1. Homologies are indicated by short lines. Dots indicate that the sequence continues but has not been determined in this particular region. The top lines show that 158 and 101 are identical at the insertion-like/F boundary. The target site known to be duplicated by F element insertion in 101 is shown boxed. The next three lines show that 30 bp are missing from the 101F sequence between 158Fa and 158Fb. Further, the sequence to the left of boundaries 4 and 6 shows the oligo(A) stretch and polyadenylation signal (underlined) at the right end of G. The bottom three lines identify boundary 3 by comparison with 5. Two short direct repeats are overlined. Sequencing was done by the method of Maxam and Gilbert (22).

somal insertion-like elements at the same nucleotide (boundary 1). Second, the analysis of two of three G/F boundaries (boundaries 4 and 6) reveals the characteristic presence of a stretch of 25 A residues at the 3' end of one strand of G. A polyadenylation site (AATAAA) precedes this run of A, in the same way as at the right terminus of every F element analyzed (3). G sequences thus join the class of oligo(A)-terminated DNA elements. Third, 30 bp present in 101F are missing from 158F, being absent at either side of the Fa or Fb regions (boundaries 2, 3 and 5). Fourth, the precise nature of the duplications of G and Fb in pDMI 158 is confirmed by the analysis of boundaries 4 and 6. In both cases, a much longer sequence has been determined than is shown, and no differences have been found between these duplicated segments of DNA. Fifth, the comparison of boundaries 3 and 5 with each
other and with pDmI 101 shows that G3 is precisely homologous to G1 up to its left end. G3 is thus a truncated version of G1 without any new sequences being inserted between the G1-homologous segment and Fb. While the left boundary of G2 has not been sequenced, detailed restriction mapping indicates that G2 is highly similar and probably identical to G3.

DISCUSSION

The structural arrangement of repeated DNA elements, which are frequently mobile, is of interest as a record of the evolutionary history of genomic rearrangements. Such information further suggests mechanisms that could be responsible for sequence rearrangements. In the present paper we report the structure of a DNA region from D. melanogaster which represents a nested organization of three otherwise distinct sequence elements. One of these elements is known to be transposable, and the same is very likely for the other two. The first element included in the region cloned in pDmI 158 is a ribosomal insertion-like sequence. Such elements are homologous to type 1 insertions which interrupt many rDNA repeating units, and are located predominantly in the chromocenter as distinct from the nucleolus organizer (1, 13, 14, 15). The suggestion that these sequences are (or have been) mobile comes from the finding of characteristic target site duplications at the ends of some ribosomal type 1 insertions, and the presence of short ribosomal DNA sequences at the termini of insertion-like sequences not associated with the rDNA locus (16, 17, 18). Insertion-like elements often occur as tandemly repeated 5 kb units and are sometimes interrupted by unrelated sequences (1, 15), e.g., the F element.

A member of the F family of sequences comprises the second unit that makes up the DNA region in pDmI 158. F elements comprise a family of repeated units of a predominant length of 4.7 kb that are dispersed in the chromocenter and at about 25 euchromatic sites, lack terminal repeats, carry an oligo(A) segment at one end, and are transposable in the genome (1, 2, 3). Five F elements analyzed in detail are colinear although two are truncated at the end opposite from the oligo(A) stretch. The insertion-like sequences in 101 and 158 are interrupted by F elements at the same position, and the left ends of these elements are also identical (Fig. 3, boundary 1). Insertion-like sequences and F sequences seem to be similarly interspersed in other distinct clones in our collection. Restriction fragments crossing the left boundaries between insertion-like and F sequences in these clones and pDmI 101 and 158 have similar size. These observations suggest that an F element
might have been introduced into an insertion-like sequence in a single event, and this composite structure would have later expanded and diverged. Alternatively, the introduction of F elements into different insertion-like elements might have been independent, in which case a high degree of site specificity for F insertion must be postulated, as suggested by the analysis of some other F targets (3).

While most F elements that have been analyzed are continuous, 158F is interrupted by a different sequence that we have called G. G sequences represent the third unit comprising the 158 structure. While these sequences have not been studied in detail we know that G is repeated in the genome in addition to its representation in the 158 region, but its copy number is quite low (Fig. 2). The most distinctive feature of G is the presence of a stretch of A residues at one end. Boundaries 4 and 6 (Fig. 3) carry an identical stretch of 23 A's preceded in each case by a polyadenylation signal (AATAAA). This property provides a close parallel to the structure of F elements, and is further reminiscent of processed pseudogenes and certain repeated sequences of mammals (7-12). Since F elements have clearly been shown to be transposable (1, 2, 3) it appears likely that G sequences are transposable as well, but no direct evidence is available on this point. Thirty bp of 158F are deleted at the original F/G boundary (Fig. 3). Therefore, it is not possible to determine whether target site duplications are induced by the insertion of G sequences into 158F.

The nested interspersion of the three units in pDMI 158 is further complicated by the tandem repetition of a unit of about 3.5 kb comprising a large part of G and a smaller part of F. The origin of the entire structure is most easily explained by assuming consecutive insertion events, first of F into the insertion-like sequence, then of G into F, followed by the expansion of the 3.5 kb G/F segment. The interspersion of otherwise distinguishable units of repetitive elements in larger complexes has been reported in several organisms. Such arrangements were suggested by Wensink and coworkers (19) for various repeated sequences in Drosophila, and clear examples have been provided for repeated sequences in the chicken (20) and the African green monkey (21). It appears that repeated sequences in the genomes of higher eukaryotes are capable of complex interspersions, at least some of which may be due to the transposable nature of many of these sequences. However, it should be remembered that not all repeated elements are subject to scrambling: members of the copia-like families of transposable sequences in Drosophila (4, 5) and the 1723 family of long repeats
in Xenopus (B. K. Kay and I. B. Dawid, unpublished) are generally found as discrete units. Thus it is clear that both distinct units and scrambled clusters contribute importantly to the repeated DNA complement of eukaryotes.

ACKNOWLEDGEMENTS

We thank Mary Thorne for preparation of the manuscript. Pier Paolo Di Nocera is on leave-of-absence from the Institute of General Pathology of the Second Medical School of the University of Naples, Italy.

REFERENCES

1. Dawid, I. B., Long, E. O., Di Nocera, P. P. and Pardue, M. L. (1981) Cell 25, 399-408.
2. Pardue, M. L. and Dawid, I. B. (1981) Chromosoma 82, 471-492.
3. Di Nocera, P. P., Digan, M. E. and Dawid, I. B. (1983) J. Mol. Biol., in press.
4. Spradling, A. C. and Rubin, G. M. (1981) Ann. Rev. Genet. 15, 219-264.
5. Scherer, G., Tschudi, C., Perera, J., Delius, H. and Pirrotta, V. (1982) J. Mol. Biol. 157, 435-451.
6. Potter, S. S. (1982) Nature (London) 297, 201-204.
7. Hollis, G. F., Hieter, P. A., McBride-V. O., Swan, D. and Leder, P. (1982) Nature (London) 296, 321-325.
8. Wilde, C. D., Crowther, P. A., Cripe, T. P., Lee, M. G. and Cowan, N. J. (1982) Nature (London) 297, 83-84.
9. Karin, M. and Richards, I. (1982) Nature (London), 299, 797-802.
10. Lemischka, I. R. and Sharp, P. A. (1982) Nature 300, 330-335.
11. Jelinek, W. R. and Schmid, C. W. (1982) Ann. Rev. Biochem. 51, 813-844.
12. Grimaldi, G. and Singer, M. F. (1982) Proc. Natl. Acad. Sci. USA 79, 1497-1500.
13. Dawid, I. B. and Botchan, P. (1977) Proc. Natl. Acad. Sci. USA 74, 4233-4237.
14. Dawid, I. B., Lauth, M. and Wellauer, P. K. (1980) In Mobilization and Reassembly of Genetic Information. W. R. Scott, R. Werner and D. R. Joseph, eds., Miami Winter Symp. 17, 217-234.
15. Kidd, S. J. and Glover, D. M. (1980) Cell 19, 103-119.
16. Dawid, I. B. and Rebbert, M. L. (1981) Nucleic Acids Res. 9, 5011-5020.
17. Roith, H., Miller, J. R., Woods, L. C. and Glover, D. M. (1981) Nature (London) 290, 749-753.
18. Roith, H. and Glover, D. M. (1981) Nucleic Acids Res. 9, 5521-5532.
19. Wensink, P. C., Tabata, S. and Pachl, C. (1979) Cell 18, 1231-1246.
20. Musti, A. M., Sobieski, D., Chen, B. B. and Eden, F. (1981) Biochemistry 20, 2989-2999.
21. Grimaldi, G. and Singer, M. F. (1983) Nucleic Acids Res. 11, 321-338.
22. Maxam, A. M. and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.