The transposable element Mdg3 in Drosophila melanogaster is flanked with the perfect direct and mismatched inverted repeats

A.A.Bayev,Jr., A.S.Krayev, N.V.Lyubomirskaya, Y.V.Ilyin, K.G.Skryabin and G.P.Georgiev

Institute of Molecular Biology, USSR Academy of Sciences, Vavilov street 32, Moscow B-334 USSR

Received 24 June 1980

ABSTRACT

Mdg3 is a family of mobile dispersed genetic elements represented by 15 copies in the haploid genome of D. melanogaster and flanked, like other similar elements, by the regions of homology. In the present work, these regions of mdg3 have been sequenced. The existence of perfect direct repeats 268 base pairs long has been demonstrated. Inverted repeats are located on the gene distal side of them. It is possible to construct a perfect 8 b.p. palindrome or a slightly mismatched 18 b.p. palindrome. The inverted repeats are flanked by two short 5 b.p. direct repeats.

INTRODUCTION

The existence of movable genetic elements in the eukaryotic genome was known for many years /1/. However, their extensive investigation became possible only after recombinant DNA techniques had been developed. Several families of mobile genetic elements dispersed throughout the whole genome (mobile dispersed genes, or mdg) were discovered in the genome of D. melanogaster /2-7/. Recently, the structural organization and transcription patterns of one of them, mdg3, were studied in detail /8, 9/. Mdg3, ~5.5 kb long, is represented by ~15 copies in the fly genome and by 200-250 copies in the genome of culture cells. Like other mdg elements /3/ mdg3 is framed by two regions of sequence homology /9/. The latter may be involved in the process of transposition and therefore the knowledge of their structure is important for understanding the transposition mechanisms in eukaryotes.

In this work, the regions of homology at the ends of mdg3 have been sequenced. It has been shown that they are represent-

ed by two perfect direct repeats 268 b.p. long. In addition, two short inverted and direct repeats have been detected at the gene-distal sides of the above mentioned regions.

METHODS

Isolation of plasmid p38m. The original hybrid plasmid p38 /9/ containing the Dm38 D. melanogaster DNA fragment with the regions of homology was too large and inconvenient for sequencing. Therefore, the insert isolated from p38 was cleaved by a mixture of BamHI and EcoRI, and fragments 2.8 kb long were isolated and ligated to pBR322 DNA cleaved by a mixture of EcoRI and BamHI. Ligation, transformation and colony hybridization were performed as described previously /7/. For colony hybridization, the nick-translated subfragments of Dm38 plasmid containing the regions of interest (see Fig. 1) were used as probes. Among the new plasmids thus obtained, a plasmid designated as p38m (plasmid N38 mini) was shown to contain only two terminal fragments carrying the sites of homology) (see Fig. 1). This was used in further experiments.

The sequencing procedure. The method of Maxam and Gilbert /10/ modified as described previously /11/ was used for sequencing. 120 µg of p38m DNA was digested with HindIII and then treated with alkaline phosphatase (BAPF, Whorthington) at pH 8.6. DNA was deproteinized, precipitated with phenol, dissolved in 40 µl of 10 mM Tris·HCl, 1 mM EDTA, pH 8.0, heated for 2 min at 72°C, and incubated at 37°C for 30 min with 250 pmoles of γ-^{32}P-ATP (1000 Ci/mmole, Amersham) and 20 units of T4-polynucleotide kinase. After gel filtration through Sephadex G-50sF, the labeled DNA was cleaved with EcoRI. The resulting fragments were separated by electrophoresis in 4% polyacrylamide gel, eluted and used for sequencing. One of these fragments (0.45 kb) which still contained the label at both ends was eluted additionally, cleaved by BspI, and the subfragments were separated by electrophoresis in 5% polyacrylamide gel.

To confirm the sequencing data for the L region (see below), a EcoRI fragment 2.8 kb long was isolated from 2.8 %

polyacrylamide gel, and cleaved with Sau3A. The resultant fragment 0.8 kb long (the longest in the digest) was isolated, labeled, cleaved with Bsp and the two fragments thus produced were sequenced. This allowed us to obtain a more easily readable sequence ladder for the junction of direct and inverted repeats (see below and Fig. 3).

After chemical cleavage, each series of products was fractionated by electrophoresis in thin 40 cm 20% gel and in 80 cm 8% gel /22/.

RESULTS

The restriction map of mdg3 cloned in Dm38 plasmid is presented in Fig. 1 (a,b). Almost the whole mdg3 is located within the HindIII fragment A. The terminal regions of sequence homology are arranged around the HindIII sites /9/.

For their sequencing, a new plasmid has been constructed which still contained the EcoRI and EcoRI-BamHI fragments of

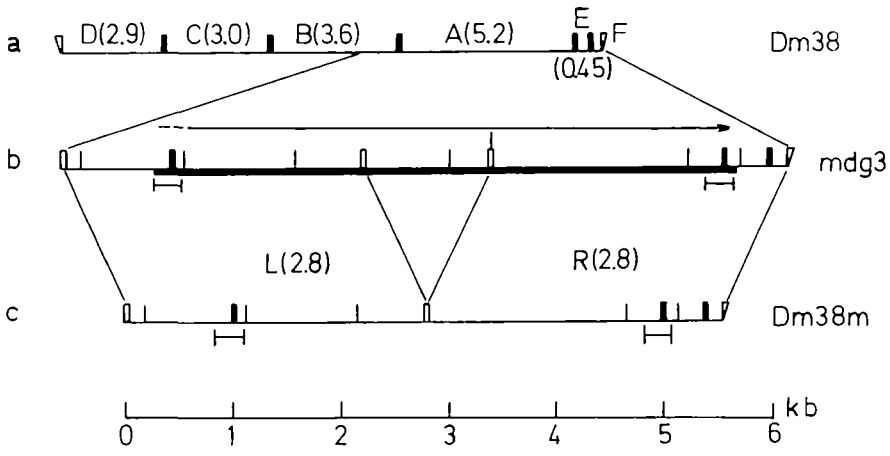

Fig. 1. The restriction maps of Dm38 and Dm38m *D. melanogaster* fragments cloned in pBR322.

(a) The map of original Dm38 cloned DNA.
(b) The enlarged part of Dm38 DNA containing mdg3.
(c) The map of recloned Dm38m DNA of *D. melanogaster*.
Restriction sites: BamHI, HindIII, EcoRI, BspI. Regions of homology ⊢⊣. Transcribed regions ▬. The main direction of transcription ⟶ /9/.

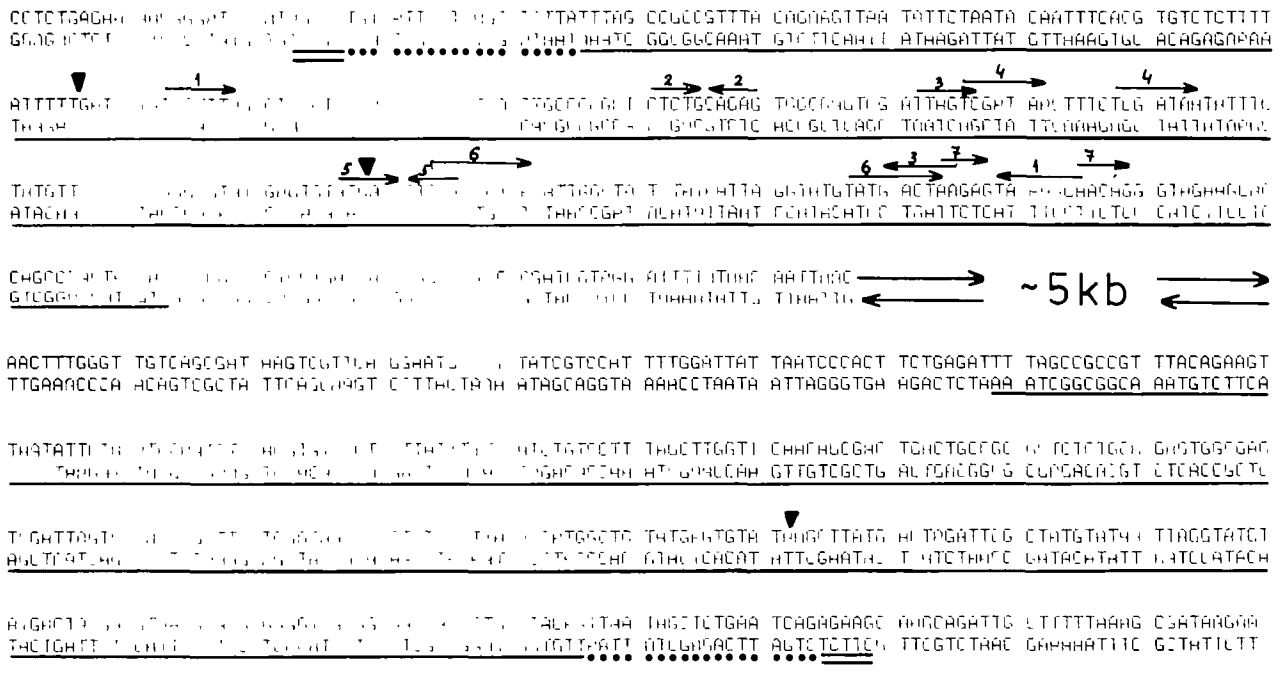

Fig. 2. The sequence of the regions L and R of mdg3.

The ⇄ -sequences (268 b.p.) are underlined with one line, the inverted repeats (18 b.p.) with dotted line and short direct repeats (5 b.p.) with two lines. Internal palindromes or direct repeats are marked with arrows (in L only). The symbol ▼ indicates HindIII and Sau3a cleavage sites that were labeled for the DNA sequence analysis.

Dm38 DNA in pBR322, but the internal EcoRI fragment was deleted (Fig. 1c).

The sequences were read by the Maxam and Gilbert technique in both directions from all of the three HindIII sites as described in Methods. Those located around the first and second HindIII restriction sites designated as L (left) and R (right) sequences cover both regions of homology detected in the hybridization experiments (compare with Fig. 1). They are shown in Figs. 2 and 3.

Both L and R regions contain exactly the same sequence of 268 b.p. in length which is oriented in the same direction. Thus, mdg3 is framed by perfect direct repeats which we designate as ξ -sequences (Fig. 2). On the gene distal sides of ξ -sequences, one can find two sequences 18 b.p. long which may be regarded as a mismatched palindrome (Fig. 3). The15 out of 18 nucleotides can be paired. The longest continuous perfect inverted repeats consist of eight base pairs. In the L-

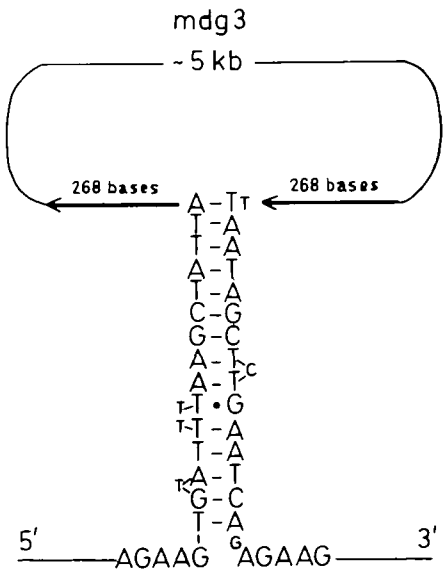

Fig. 3. The demonstration of mismatched inverted repeats on the mdg3-distal sides of ξ -sequences.

Normal base pairs are separated by — , GT hydrogen pair by • .

-region, such a repeat immediately precedes the ξ -sequence whereas, in the R-region, it is separated from the ξ by one base pair. In their turn,these 18 b.p. nonperfect repeats are flanked with short (5 b.p.) direct repeats (Figs. 2, 3).

Fig. 4 demonstrates the structural gels covering the regions on the gene distal sides of long direct repeats.

No homology could be detected beyond the described region in any direction.

We also have sequenced about 200 base pairs in the region around the third HindIII restriction site. The sequenced area is located about 150 b.p. apart from the R-region sequence. Thus, it does not belong to mdg3. However, we present this sequence in Fig. 5 to demonstrate the existence of a peculiar, extremely (A+T)-rich sequence in the genome of D. melanogaster not far from the target site for mdg3 insertion.

DISCUSSION

In this work, we have sequenced the regions of homology located at the ends of a mobile dispersed genetic elements in D. melanogaster. The existence of direct repeats at the termini of mdg3 suggested by previous hybridization experiments /9/ has been proved by sequencing. The direct repeat consisting of 268 base pairs are perfect. No mismatched nucleotides have been found. The absence of deviations indicates that either the perfectness of repeated sequences is under the selection pressure or the direct repeats in a particular location of mdg3 were formed rather recently, possibly in the course of a recent transposition event.

The sequenced copy of the mdg3 family originates from 93C region of D. melanogaster chromosomes. This is the only region to which the HindIII fragments C and D of p38 DNA (see fig. 1) do hybridize /9/. However, only a few of individuals analyzed with the aid of in situ hybridization have been found to contain mdg3 at this particular site of a chromosome. In most cases, 93C region does not hybridize to an mdg3 probe. It would be interesting to analyze the ξ -sequences of mdg3 copies originating from the other sites of chromosomes, in particular from the sites where mdg3 is usually located.

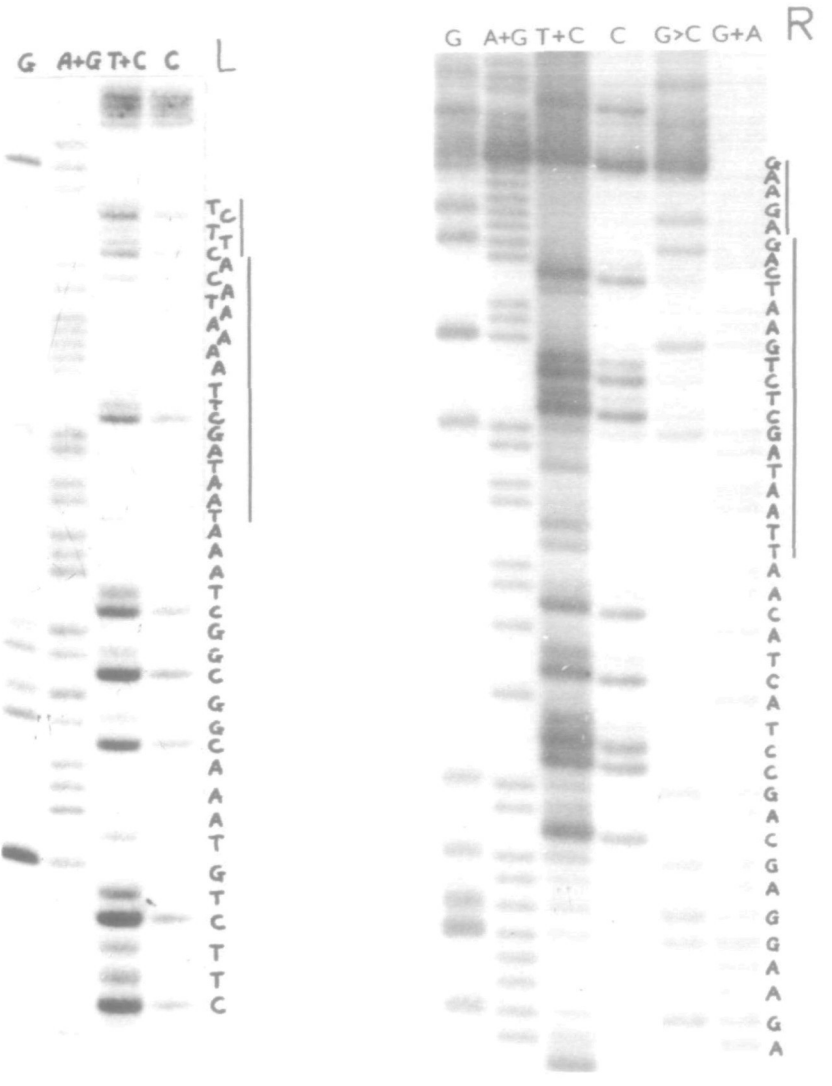

Fig. 4. The sequencing gels demonstrating the existence of short inverted and direct repeats at the ends of mag3.

The lines G, A+G, T+C, and C indicate chemical reactions according to /10, 11/, and G,>C, G+A are those employing diethylpyrocarbonate (A.Krayev, in preparation). L-sequence flanking the left ⊱ was read from the right to the left. R-sequence flanking the right ⊱ was read from the left to the right.
The 5 nucleotide direct repeats and 18 nucleotide inverted repeats are marked by vertical lines.

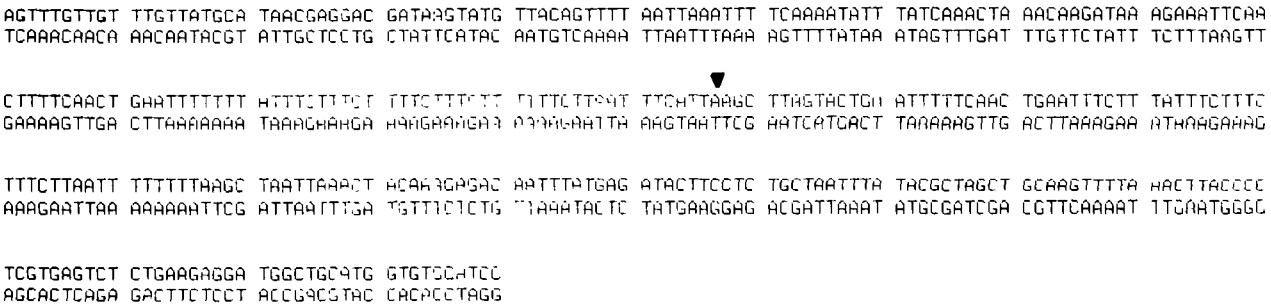

Fig. 5. The AT-rich sequence around the third HindIII site located outside the mag3 region.

The symbol ▼ indicates HindIII cleavage site that was labeled for the DNA sequence analysis.

The analysis of the ξ -sequence reveals several short palindromes there as well as internal direct repeats (5–8 base pairs in length) (see Fig. 2). Their significance remains unclear.

Previously, we found that the transcription of mdg3 was terminated and possibly started within the ξ -sequence /8, 9/. Also the quasi-symmetric character of transcription was demonstrated. Analysis of the ξ -sequences allows one to find several sites similar to the Hogness boxes and to the known termination signals /12/ in both orientations. However, it is impossible to deduce the real localization of the transcription initiation and termination sites from the DNA sequence data only. Additional experiments are necessary and these are in progress now.

The most interesting observation is concerned with the organization of gene distal termini in the regions of homology. We found there mismatched inverted repeats 18 b.p. long flanked with perfect direct repeats 5 b.p. long. Such a structure (a non-perfect palindrome and direct repeats 5 or 9 b.p.long) is typical of the ends of bacterial insertion sequences (IS) and transposones /13-18/. The short direct repeats there originate from the duplication of the genomic target sequence into which a transposone is inserted. By analogy, we may suggest the same origin for our 5 b.p. repeats but such a possibility should be directly tested by sequencing the ends in other members of mdg3 family, as well as by studying the target sequence in flies without an mdg3 insertion in 93C region.

In bacterial cells, the mismatched inverted repeats are parts of transposable elements /13, 14, 16, 18/. Their origin in D. melanogaster is under investigation now.

It should be pointed out that the existence of long direct repeats at the termini of an inserted sequence is typical of endogenous pro-retroviruses /19/. However, no details are available so far about the sequence organization at the junction sites of the pro-retroviruses and the host genome. The presence of similar long direct repeats in the mobile dispersed genes of D. melanogaster /3, 20/ and yeast /21/ may indicate the close relationship between these two types of genetic

elements.

ACKNOWLEDGEMENTS

The authors are grateful to Prof. A.A.Bayev for valuable discussions and to Drs. A.Janulaitis, A.Bocharov and P.Rubtsov for restriction enzyme preparations.

REFERENCES

1 McClintock, B. (1957) Cold Spring Harbor Symp. Quant. Biol. 21, 197-216.
2 Ilyin, Y.V., Tchurikov, N.A., Ananiev, E.V., Kyskov, A.P., Yenikolopov, G.N., Limborska, S.A., Maleeva, N.E., Gvozdev, V.A., and Georgiev, G.P. (1978) Cold Spring Harbor Symp. Quant. Biol. 24, 959-969.
3 Finnegan, D.J., Rubin, G.M., Young, M.V., and Hogness, D.S. (1978) Cold Spring Harbor Symp. Quant. Biol. 42, 1053-1063.
4 Ananiev, E.V., Gvozdev, V.A., Ilyin, Y.V., Tchurikov, N.A., and Georgiev, G.P. (1978) Chromosoma 70, 1-17.
5 Georgiev, G.P., Ilyin, Y.V., Kyskov, A.P., Tchurikov, N.A., Yenikolopov, G.N., Gvozdev, V.A., and Ananiev, E.V. (1977) Science 195, 394-397.
6 Tchurikov, N.A., Ilyin, Y.V., Ananiev, E.V., and Georgiev, G.P. (1978) Nucleic Acids Res. 5, 2169-2187.
7 Tchurikov, N.A., Zelentsova, E.S., and Georgiev, G.P. (1980) Nucleic Acids Res. 8, 1243-1258.
8 Ilyin Y.V., Chmeliauskaite, V.G., and Georgiev, G.P. Nucleic Acids Res., (1980) 8, 3439-3458
9 Ilyin, Y.V., Chmeliauskaite, V.G., Ananiev, E.V., and Georgiev, G.P. (1980) Chromosoma, in press.
10 Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560 564.
11 Skryabin, K.G., Zaknaryev, V.M., and Bayev, A.A. (1978) Proc. Acad. Sci. (USSR) 241, 488-491.
12 Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
13 Calos, M.P., Johnsrud, L., Miller, J.H. (1978) Cell 13, 411-418.
14 Grindley, N.D.F. (1978) Cell 13, 419-426.
15 Heffron, F., McCarthy, B.J., Ohtsubo, H., and Ohtsubo, E. (1979) Cell 18, 1153-1164.
16 Ghosal, D., Sommer, H. and Saedler, H. (1979) Nucleic Acids Res. 6, 1111-1122.
17 Kleckner, N. (1979) Cell 16, 711-720.
18 Reed, R.R., Young, R.A., Steitz, J.A., Grindley, N.D.F., and Guyer, M.S. (1979) Proc. Natl. Acad. Sci. USA 76, 4882-4886.
19 Gilboa, E., Mitra, S.W., Goff, S. and Baltimore, D. (1979) Cell 18, 93-100.
20 Ilyin, Y.V., Chmeliauskaite, V.G., Ananiev, E.V., Lyubomirskaya, N.V., Kulguskin V.V., Bayev, A.A., Jr., and Georgiev, G.P. Nucleic Acids Res., submitted.

21 Philippsen, P., Gafner, J., Eibel, H., Brennan, M., Stotz, A., and Hohn, B. (1979) Hoppe-Seyler's Z.Physiol.Chem. 360, 1038.

22 Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.