Volume 9 Number 21 1981 Nucleic Acids Research

Duplicated rDNA sequences of variable lengths flanking the short type I insertions in the rDNA of Drosophila melanogaster

Heli Roiha and David M.Glover

Cancer Research Campaign, Eukaryotic Molecular Genetics Research Group, Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK

Received 18 September 1981

ABSTRACT

We describe cloned segments of rDNA that contain short type I insertions of differing lengths. These insertions represent a coterminal subset of sequences from the right hand side of the major 5kb type I insertion. Three of these shorter insertions are flanked on both sides by a short sequence present as a single copy in uninterrupted rDNA units. The duplicated segment is 7, 14 and 15 nucleotides in the different clones. In this respect, the insertions differ from the 5kb type I insertion, where the corresponding sequence is found only at the right hand junction and where at the left hand side there is a deletion of 9 nucleotides of rDNA (Roiha et al.,1981). One clone is unusual in that it contains two type I insertions, one of which is flanked by a 14 nucleotide repeat. The left hand junction of the second insertion occurs 380 nucleotides downstream in the rDNA unit from the first. It has an identical right hand junction to the other elements and the 380 nucleotide rDNA sequence is repeated on both sides of the insertion. We discuss the variety of sequence rearrangements of the rDNA which flank type I insertions.

INTRODUCTION

The rDNA of Drosophila melanogaster can contain two types of non-homologous insertion sequences in the 28S gene (1-8). These insertions are unlike intervening sequences found in other genes and appear to have a repressive effect on the transcription of rDNA. It is possible to detect only a few high molecular weight transcripts containing insertion sequences and yet the nucleotide sequence of the region in which rDNA transcription initiates is identical in units containing insertions and those without (9,10).

The type I insertions are found in more than 50% of the rDNA units of the X chromosome and not in the rDNA of the Y chromosome (11,12,13). The most common type I insertion is about 5kb long, but in addition shorter insertions have been described which are mainly comprised of 0.5kb and 1kb of DNA from the right hand side of the 5kb sequence. Roughly 50% of the type I sequences are found at other chromosomal sites, where they occur predominantly in

tandem arrays (14). These type I sequences have probably undergone transposition from the nucleolus since the units within the tandem arrays are flanked by very short segments of the 28S gene (15). Many Dipteran flies have insertion sequences in their rDNA at the same site as D.melanogaster, as determined by low resolution mapping techniques (16-19). In the case of D.virilis the insertions have been shown by sequencing experiments to be at exactly the same site as the type I sequences of D.melanogaster. The D.virilis insertions, however, are flanked on both sides by a short segment of the gene present only once in uninterrupted units (20). In our previous sequence analysis of the junctions of several long type I insertions with the D.melanogaster 28S gene, we found only a single copy of this rDNA sequence to the right of the insertion. At the left hand junction nine nucleotides of rDNA are deleted (15).

In order to determine whether this was the only arrangement of sequences flanking type I insertions in the genome we decided to sequence the junctions between the short type I insertions and rDNA. In the cloned EcoRI fragments of rDNA which we examine in this paper, the short type I insertions are flanked by a duplicated sequence of 7 to 15 nucleotides with no deletions of flanking rDNA.

MATERIALS AND METHODS

Construction of Cloned DNAs. Recombinant DNAs were constructed in collaboration with Simon Kidd and Mick Browne. D.melanogaster embryonic DNA was digested to completion with EcoRI and enriched for fragments of 12kb-17kb, either by preparative sucrose gradient sedimentation or preparative electrophoresis on agarose gels (14). Approximately 0.25ug of gradient purified DNA was ligated to 0.5ug of purified left and right arms of λgt.λWES (21) at a concentration of about 100ug/ml. The ligated DNA generated 2 x 10^6 plaque forming units in the in vitro packaging reaction (22). From 2 x 10^4 plaques which were screened, 77 were selected which hybridised to the 1kb BstI fragment and not the 4kb HindIII/BstI fragment of the Dml03 insertion. One of these phages, λMB27, is described in this paper. Approximately 1ug of the gel purified EcoRI fragments were ligated with 0.5ug of dephosphorylated EcoRI linear molecules of pAT153 (23). The ligation mix generated 7 x 10^3 ampicillin resistant transformants of E.coli. Of these, 16 hybridised with the 1kb BstI fragment and not with the 2.6kb SmaI fragment of Dml03. Two of these recombinants, RI9 and RI10, are described in this paper.

DNA Sequencing. All sequencing was performed as described by Maxam and Gilbert (24). Restriction fragments were labelled by 'filling in' the cohesive termini using reverse transcriptase (from J. Beard). The KpnI site was labelled using terminal transferase mediated addition of ^{32}P-cordycepin (Amersham).

RESULTS

We have previously examined the nucleotide sequence at the junctions of type I sequences with rDNA in three configurations (15):
- an rDNA unit containing a 5kb type I insertion (the rDNA segment Dm103, reference 25);
-an rDNA unit containing both a type I insertion and, 51 nucleotides upstream, a type II insertion (Dm207, references 5 and 8);
- a fragment from the chromocentral heterochromatin consisting of 5 tandemly arranged type I elements, each joined to a 21 segment of rDNA on its left and a 13 nucleotide segment on its right (Dm219, reference 14).
In each of these cases the junctions are identical. At the left hand side of the type I insertion there is a deletion of nine nucleotides of rDNA. The right hand junction corresponds to the site of the rDNA insertions in D.virilis.

There are in addition a set of shorter type I insertions which we had not examined in the above experiments. The majority of these are either 0.5kb or 1kb in length and have sequence homology to the extreme right hand part of the 5kb insertion. We selected clones which would hybridise to sequences from this right hand region but not to sequences from the left hand part of the long type I insertions (see Materials and Methods). We sequenced the entire insertions in three of these clones, RI9, RI10, and MB27, following the strategies shown in figure 1. The 28S rRNA of D.melanogaster consists of four polynucleotide chains (the 2S RNA, 5.8S RNA, 28Sα RNA, and 28Sβ RNA) held together by hydrogen bonding. The insertions divide the rDNA coding for 28Sβ RNA into two regions, 28Sβ1 and 28Sβ2, and it is this region of the 28S gene which is shown by the shaded blocks in figure 1. The top map shows the right hand portion of the 5kb Dm103 insertion which we have sequenced. This sequence is shown in figure 2. Each of the shorter insertions is an extremely homologous segment of this sequence, showing only 0.28% to 0.40% sequence divergence (see the legend to Fig.2). Each has an identical right hand junction with rDNA to the one found in Dm103. The positions of the left hand ends of the shorter insertions are indicated by arrows on the Dm103

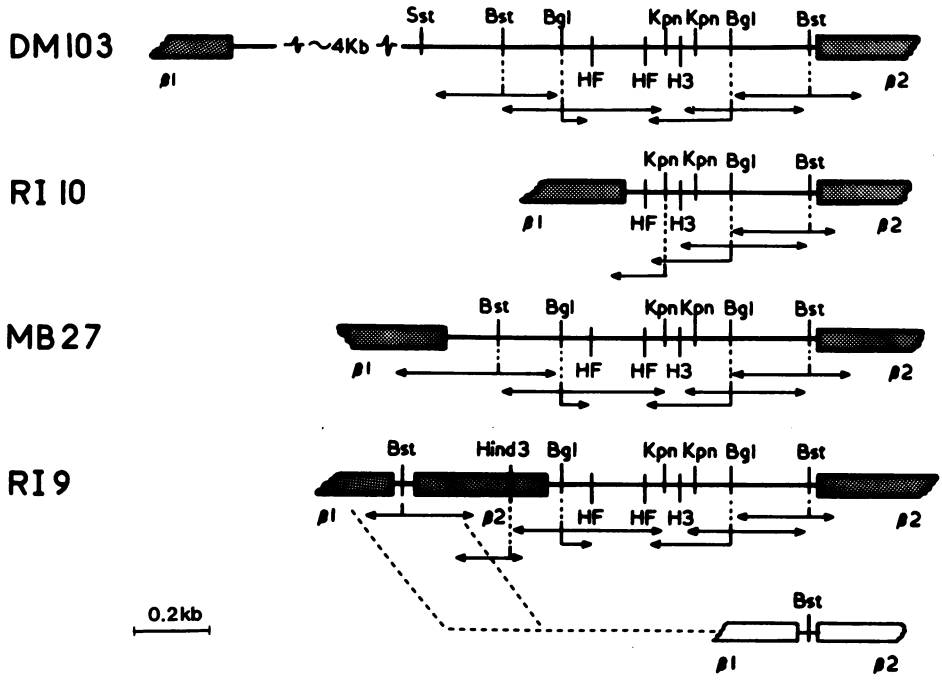

Figure 1. Sequencing Strategies
The physical maps show the regions of the cloned rDNA segments which were sequenced. Dm103 is the cloned EcoRI fragment described by Glover et al. (25). The selection of the three other cloned EcoRI fragments is described in the Materials and Methods. The arrows show the directions and extent to which sequences were read from the terminally labelled nucleotide. The shaded blocks represent rDNA and the solid lines the insertion sequences. The clone RI9 contains two insertions, which have identical right hand junctions. The dotted lines indicate the displacement of the left hand insertion to bring its right hand junction into alignment with the other cloned segments (Bgl = BglII; Bst = BstI; H3 = HaeIII; HF = HinfI; Kpn = KpnI; Sst = SstI).

sequence in figure 2. In each case there is a direct transition from insertion sequence to rDNA sequence.

We find a variety of arrangements of the rDNA sequences at the left hand junctions. Three of the insertions we have sequenced are flanked on both sides by an rDNA segment which is present as a single copy in the uninterrupted rDNA unit (Figs.3 and 4). The left end points of these three insertions are shown by the solid arrows in figure 2. The duplicated sequence varies from 7 to 15 nucleotides. In two cases it is not possible to

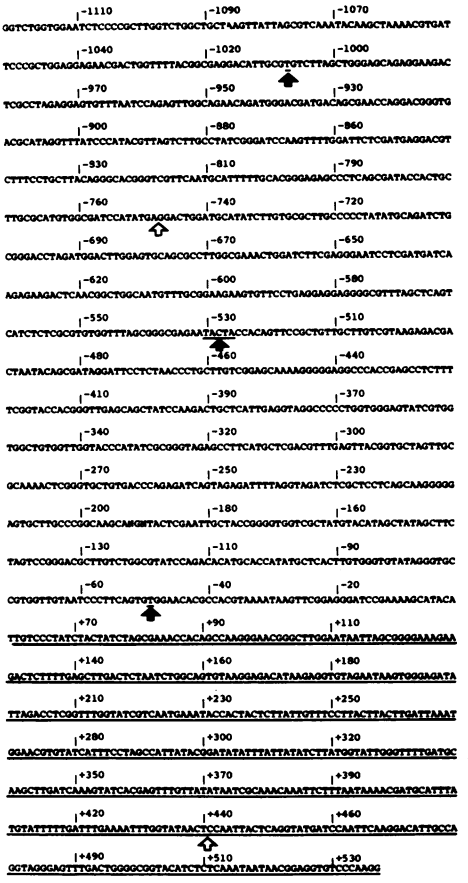

Figure 2. A Contiguous Nucleotide Sequence at the Right Hand Junction of the Dm103 Insertion with rDNA.
This sequence was principally derived from sequencing experiments illustrated on the Dm103 map in Fig.1. Numbering of the sequences follows the nomenclature of Roiha et al. (15), where the right hand side of the insertion is given negative numbers from the point at which it meets rDNA, and the rDNA is numbered with respect to the position of the type II insertion. The rDNA sequence is continuously underlined (nucleotides +61 to +537). The sequence from nucleotide +61 to +342 is from Roiha et al. (15). The rDNA sequence beyond nucleotide +342 was obtained by sequencing rightward from the single HindIII site in the 28Sβ2 segment of the gene. The solid arrows indicate the left hand end points of the shorter insertions which are flanked by duplicated sequences of 7 to 15 nucleotides (see also Fig. 3). The open arrows indicate the points at which the left hand of the 747 nucleotide insertion in RI9 joins with the 28Sβ2 gene (see also Fig. 3). The sequences of the shorter insertions are identical with that of Dm103 except at the following positions: Clone RI10: -604, C; -123, G. Clone MB27: - 933, C; -888, C; -82, C. Clone RI9 (747 nucleotide insertion) -604, C; -286, G; -264, A.

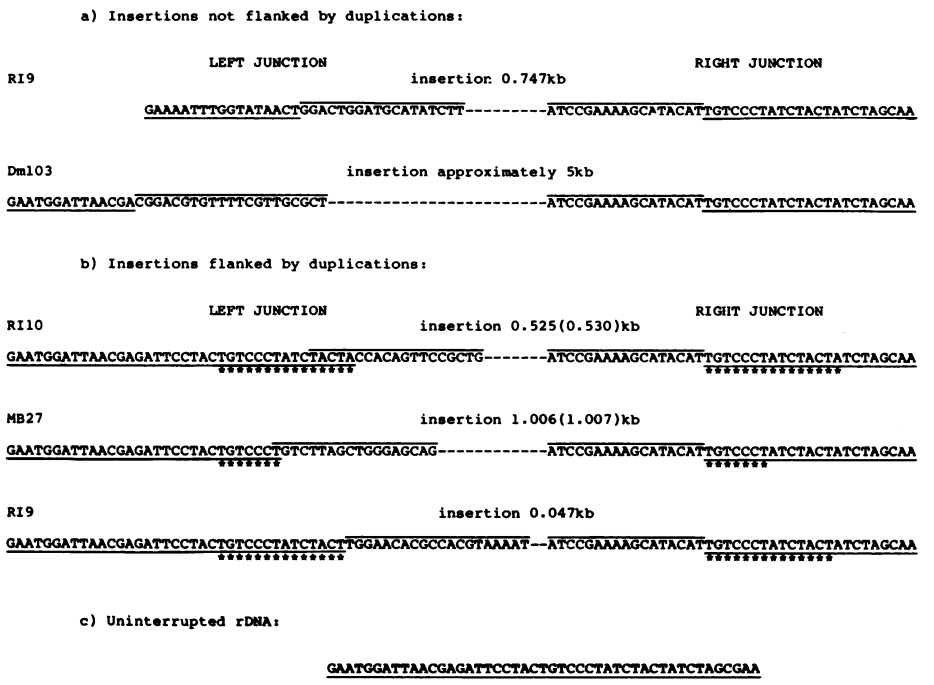

Figure 3. A Comparison of Insertion Junctions
The rDNA sequences to the left and right of insertions are shown in alignment to facilitate comparison. The one exception is the sequence to the left of the 747 nucleotide insertion in RI9, where the rDNA originates from the 28Sβ2 region rather than the 28Sβ1 region as in the other cases (see also Fig. 1). This cloned segment is unusual in that it contains two insertions 380 nucleotides apart. Strictly speaking this insertion is flanked by a duplication of this 380 nucleotide sequence, but we have placed it in a separate category together with Dml03, since the structure of the clone is so unusual. The sequence at the left hand junction in Dml03 is from Roiha et al. (15). The rDNA sequences are underlined and the insertion sequences overlined. The duplicated sequence is indicated by the asterisks. The sequence of uninterrupted rDNA (15) is shown for comparison.

say whether part of the duplication is derived from rDNA or the insertion since common nucleotides occur at these positions in the two sequences (see Fig.3). The rDNA segment RI10 contains a 0.5kb insertion flanked by a 15 nucleotide repeated sequence. The sequence repeat is entirely contained within rDNA at the right hand junction. At the left hand side, however, the junction cannot be unambiguously placed; 10 nucleotides must correspond to rDNA sequence, but the remaining 5 could correspond to sequences contiguous

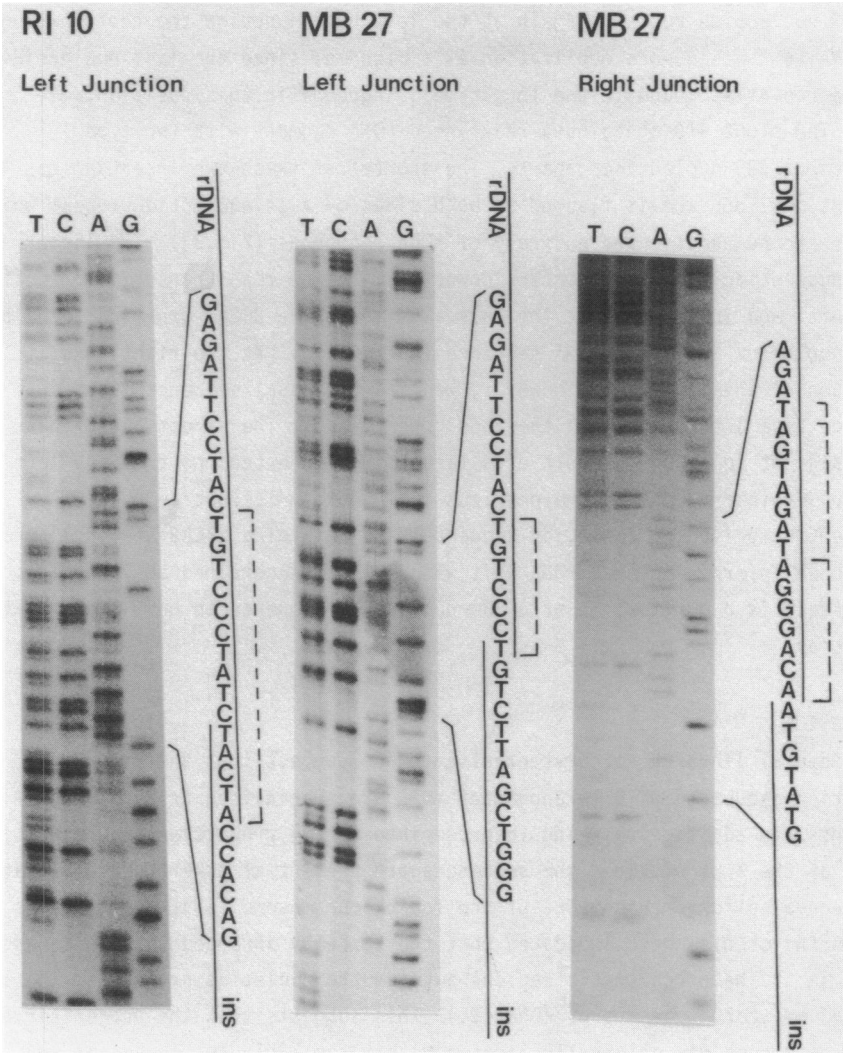

Figure 4. Sequencing Gels Showing the Left Hand Junctions of the Insertions in RI10 and MB27 and the Right Hand Junction in MB27.
The fragment of RI10 which contains the left junction was generated by KpnI and BglII cleavage, and was 3' labelled at the KpnI site. The MB27 fragments were generated by cleavage with BstI and BglII (left junction) or BstI and EcoRI (right junction) and were both labelled at their BstI sites. The rDNA and insertion sequences are indicated by solid lines placed against the sequence. The dashed line indicates the segment of rDNA duplicated on either side of the insertion. The variations seen in the length of the duplication are indicated against the sequence of the MB27 right hand junction. This junction is identical for each clone.

with either rDNA or the insertion. The clone MB27 has a 1kb insertion flanked by a 7 nucleotide repeat. Again at the left hand junction the contribution of rDNA to the sequence duplication is ambiguous, since the last nucleotide of the repeat is found in the long type I sequence in this position.

The clone RI9 contains a rearranged rDNA segment with two type I insertions 380 nucleotides apart. The shorter of these two insertions is 47 nucleotides long and is flanked on both sides by a 14 nucleotide repeat which appears to be constituted entirely of rDNA sequences (Fig.3). The larger (747 nucleotide) insertion occurs downstream in the rDNA transcription unit. Its left hand junction is at the unusual site in the 28Sβ2 gene indicated by the open arrow above the rDNA sequence in figure 2. At the right hand junction of this insertion, however, we find the usual sequence junction between type I sequences and the 28Sβ gene (Fig.3). The structure of this rDNA segment could be a result of a large deletion extending from the left of the longer insertion into the previous rDNA unit. Alternatively, a DNA fragment comprised of insertion sequences and adjoining 28Sβ2 sequences could have been inserted into an rDNA unit which already contained an insertion. The effect is a duplication of a 380 nucleotide sequence on both sides of the insertion.

DISCUSSION

Several lines of circumstantial evidence suggest that the type I insertion sequences of D.melanogaster rDNA are specialised transposable elements. In addition to being inserted into a high proportion of the 28S genes of the X chromosome, the sequences are present at other heterochromatic loci and also in a region 102C of the fourth chromosome (13). Our previous sequencing studies have indicated that the tandemly arranged type I sequences found in the heterochromatic regions outside the nucleolus organiser (14) are flanked by short segments of rDNA (15). This suggests that the progenitor of these sequences was originally located within the nucleolus organiser and following excision became tandemly amplified either by autonomous replication or by unequal exchange following its insertion at a new chromosomal site. The rDNA segments which flank these type I units are identical to those immediately adjacent to the type I insertions in the cloned EcoRI fragments of rDNA, Dm103 and Dm207 (15). The type I insertion in Dm103 is approximately 5kb in length and is representative of about 60% of the rDNA units on the X chromosome. The rDNA segment Dm207 contains both a type I and a type II insertion. The identity of the sequence junctions in all these

cases suggests that they represent a principal sequence arrangement in the genome (15). This is surprising since to the left of the insertion there is a deletion of nine residues of rDNA. Thus it would seem that an rDNA unit which is incapable of producing functional rRNA has been spread around the genome.

In this paper we characterise rDNA units containing type I insertions which differ from the 5kb insertion in Dm103 in several respects. These units have shorter insertions which are flanked on both sides by a nucleotide sequence present only once at this site in uninterrupted rDNA units and which do not have deletions in the flanking rDNA sequence. The 5kb and 10kb insertion sequences in the rDNA of D.virilis occur at this same site and are also flanked by such a duplicated sequence (20). We have previously referred to this duplicated sequence as the β* segment when making comparisons with D.melanogaster rDNA (15). It corresponds to the small segment of rDNA found on the right hand side of the type I elements in Dm219. In this paper we show that the length of the duplicated sequence varies, and so for the purposes of the following discussion we re-define the β* segment as the rDNA sequence which is duplicated. The occurence of insertion elements at identical sites in the rDNA of two Drosophila species could indicate that insertion sequences have been stably maintained at this site, just as the positions of intervening sequences have been conserved in genes which encode polypeptides. However, several pieces of evidence suggest that transpositions of the type I sequences have been ongoing since the divergence of the D.melanogaster and D.virilis species. The rDNA sequence in this region is strongly conserved throughout eukaryotes. Although the insertions of D.virilis and D.melanogaster share some homology, thermal denaturation of heteroduplexes formed between the insertions indicates extensive mismatching (16). In contrast, the chromocentral type I sequences of D.melanogaster show obvious identity with the rDNA insertions. Indeed, the junctions of 21 and 13 nucleotide rDNA segments with chromocentral type I units of Dm219 are identical to the junctions of these sequences in the nucleolar rDNA (15). This suggests that these insertion sequences together with segments of flanking rDNA have migrated from the nucleolus relatively recently. Furthermore, the shorter type I insertions form a coterminal subset of sequences from the 5kb insertion of Dm103, suggesting that they have been independently derived from these longer insertion elements.

Since the insertions are always found at the same site, it is likely that transposition could occur by a process of site specific recombination

analogous to that utilised by the bacteriophage λ genome for integration into and excision from the E.coli chromosome. The duplicated sequence would corrrespond to the duplicated att site sequence found on either side of the bacteriophage genome in lysogens. rDNA insertion sequences flanked by duplicated sequences could be excised by homologous recombination in a "Campbell-type" process. This would generate an uninterrupted rDNA unit and an extrachromosomal circle of the insertion sequence, each containing a single copy of the β* element. This extrachromosomal element would then be capable of integration at the same site in other "empty" rDNA units. The strong conservation of the sequence arrangement at the right hand junction of the type I insertion in all the clones which we have analysed (see Figs.2 and 3, and reference 15) suggests that this would be the recognition site for the recombinational process. Only one copy of the β* segment at this site may be necessary to promote recombination, since the insertion element in Dm103 is flanked by a single copy of this sequence and has undergone transposition to alternative chromosomal sites.

The 0.5kb element in RI10 could arise by recombination between the sequence TACTA (nucleotides -530 to -526, Fig.2) in a 5kb insertion and its homologue (nucleotides +71 to +75, Fig.2) in rDNA. This would generate an extrachromosomal insertion element with a 15 nucleotide homology to rDNA which could therefore generate a 15 nucleotide duplication upon re-insertion. Homologous pairing need not be an absolute requirement for the generation of new arrangements of insertions. Aberrant excisions of the lambdoid phages can result in a variety of sequence rearrangements about the att site, and similar events could explain the origins of the 1kb insertion of MB27 and the 47 nucleotide insertion of RI9. At the position in the Dm103 sequence corresponding to the left hand "breakpoint" necessary to produce these shorter insertions, there is limited homology with the β* sequence (viz. nucleotides -1007 to -1004, TGTC; nucleotides -51 to -49, TGT; Fig.2). Whilst this partial homology might favour aberrant excisions at these sites, the "breakpoint" has clearly been out of register with the homology. These two insertions are flanked by duplicated sequences which differ in length. This is best explained if there is flexibility of the recombinational exchange within the β* site during the generation of shorter insertions. Once an extrachromosomal element has been generated containing an rDNA segment of a particular length, it could be perpetuated in the genome by homologous recombination into rDNA. Equally any novel rDNA unit generated in such a way could be spread throughout the nucleolus organiser by gene

conversion or unequal exchange mechanisms acting on the rDNA units per se.

If the rDNA sequence which becomes duplicated is an essential feature of the recognition site, one might expect to find rearrangements on either side of insertions where such a sequence occurs. The deletion of rDNA to the left hand side of the 5kb insertions could be the consequence of an aberrant excision. Similarly one would expect to find evidence of aberrant excisions of sequences extending from a β* element rightwards into the 28Sβ2 sequence. The duplication of a much longer part of the 28Sβ2 sequence in the clone RI9 could be explained by the re-insertion of such an aberrantly excised element into another rDNA unit containing an insertion. Alternatively, this clone could be the result of a large deletion extending from the left of the 747 nucleotide insertion into the adjoining rDNA unit.

There is abundant evidence that many sequence rearrangements have occurred in this region of the 28S gene, and from the present arrangement of sequences we infer that there is a mechanism for site specific recombination at this site. This must, however, remain speculative since direct evidence for ongoing transposition has not yet been presented. Such a mechanism could also allow insertion elements to serve as "end points" for the excision of large blocks of rDNA from the tandemly arranged ribosomal genes. The selective advantage of systems that would either allow unequal exchange of blocks of rDNA or extrachromosomal replication of rDNA have been previously extolled as a potential means for correcting deficiencies of rDNA in the phenomenon of rDNA magnification (26,27). The sequence analysis that we present in this paper suggests a mechanism whereby the rDNA insertion elements themselves or large blocks of rDNA flanked by insertion elements could participate in such a processs. It will be of interest to examine the involvement of type I sequences in the rDNA magnification process.

ACKNOWLEDGEMENTS

This work was supported by Project Grants from the Medical Research Council. DMG holds a Career Development Award from the Cancer Research Campaign. We wish to thank both Simon Kidd and Mick Browne for their valuable contributions to this work.

REFERENCES

1. Glover, D.M. and Hogness, D.S. (1977). Cell 10, 167-176.
2. White, R.L. and Hogness, D.S. (1977). Cell 10, 177-192.
3. Wellauer, P.K. and Dawid, I.B. (1977). Cell 10, 193-212.
4. Pellegrini, M., Manning, J. and Davidson, N. (1977). Cell 10, 213-224.

5. Glover, D.M. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 4932-4936.
6. Dawid, I.B., Wellauer, P.K. and Long, E.O. (1978). J. Mol. Biol. 126, 749-768.
7. Wellauer, P.K. and Dawid, I.B. (1978). J. Mol. Biol. 126, 769-782.
8. Roiha, H. and Glover, D.M. (1980). J. Mol. Biol. 140, 341-355.
9. Long, E.O. and Dawid, I.B. (1979). Nucleic Acids Res. 7, 205-216.
10. Long, E.O., Rebbert, M.L. and Dawid, I.B. (1981). Proc. Natl. Acad. Sci. U.S.A. 78, 1513-1517.
11. Tartof, K.D. and Dawid, I.B. (1976). Nature 263, 27-30.
12. Wellauer, P.K., Dawid, I.B. and Tartof, K.D. (1978). Cell 14, 269-278.
13. Peacock, W.J., Appels, R., Endow, S. and Glover, D.M. (1980). Genetical Research 37, 209-214.
14. Kidd, S.J. and Glover, D.M. (1980). Cell 19, 103-119.
15. Roiha, H., Miller, J.R., Woods, L.C. and Glover, D.M. (1981). Nature 290, 749-753.
16. Barnett, T. and Rae, P.M.M. (1979). Cell 16, 763-775.
17. Renkawitz-Pohl, R., Glatzer, K.H. and Kunz, W. (1980). Nucleic Acids Res. 8, 4593-4611.
18. Beckingham, K. and White, R.L. (1979). J. Mol. Biol. 137, 349-373.
19. Renkawitz-Pohl, R., Matsumoto, L. and Gerbi, S.A. (1981). Nucleic Acids Res. 9, 3747-3764.
20. Rae, P.M.M., Kohorn, B.D. and Wade, R.P. (1980). Nucleic Acids Res. 8, 3491-3504.
21. Leder, P., Tiemeier, D. and Enquist, L. (1977). Science 196, 175-177.
22. Hohn, B. and Murray, K. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 3259-3264.
23. Twigg, A.J. and Sherratt, D. (1980). Nature 283, 216-218.
24. Maxam, A.M. and Gilbert, W. (1980). Methods in Enzymology, K. Moldave and L. Grossman, Eds. 65, 499-560, Academic Press, New York.
25. Glover, D.M., White, R.L., Finnegan, D.J. and Hogness, D.S. (1975). Cell 5, 149-157.
26. Tartof, K.D. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 491-500.
27. Ritossa, F., Scalenghe, F., Di Turi, N. and Contini, A.M. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 483-490.