Volume 12 Number 11 1984 Nucleic Acids Research

Ribosomal DNA genes of *Bombyx mori*: a minor fraction of the repeating units contain insertions

Rena Lecanidou[1,2], Thomas H.Eickbush[3] and Fotis C.Kafatos[1,4]

[1]Department of Cellular and Developmental Biology, The Biological Laboratories, Harvard University, Cambridge, MA 02138, USA, and [2]Department of Biochemistry, Cell and Molecular Biology and Genetics, University of Athens, Panepistimiopolis, Kouponia, Athens 15701, Greece

Received 27 February 1984; Revised and Accepted 8 May 1984

ABSTRACT

We have analyzed multiple recombinant DNA clones containing ribosomal RNA repeat units of the silkmoth, Bombyx mori. In combination with genomic DNA blots, analysis of these clones indicated that the rDNA repeat of B. mori is 10.8 kilobase pair in length and tandemly repeated in the genome, as reported by Manning et al. (18). However, contrary to that report, approximately 12% of the rDNA cistrons are interrupted by insertions of non-ribosomal DNA. Two classes of DNA insertions were identified. In one class the insertions are positioned in a region of the 28S coding sequence similar to that of the predominant rDNA insertions found in a variety of Dipteran and Tetrahymena species. In the second class, probable insertions are found close to the 3' terminus of the 28S coding sequence. Restriction enzyme analysis indicates that the two classes of insertions are not related.

INTRODUCTION

The genes for rRNA are tandemly repeated in the genomes of all higher eukaryotes which have been investigated. While in most species these repeating units have been found to contain only continuous genes, in certain species non-ribosomal DNAs have been found interrupting the rRNA units (reviewed in ref. 1). Interrupted genes coding for the larger rRNA have been found in Physarum polycephalum (2,3), several Tetrahymena species (4,5), yeast mitochondrial DNA (6), Chlamydomonas reinhardi chloroplast DNA (7) and in six out of seven dipteran insect species studied (8-14). In all cases except diptera, all rDNA genes in the species are interrupted and thus must be transcribed, the interruptions being characterized as introns. In the diptera, only a fraction of the ribosomal genes are interrupted, and these copies apparently are not transcribed (15; see review by Beckingham, ref 16); in that case it seems more appropriate to

call these interruptions insertions rather than introns.

B. mori has approximately 240 copies of rRNA genes (17) which are organized in 10.8 kilobase pair tandem repeats (18). Contrary to a previous report (18), we have found that insertions are present in a fraction of the rRNA genes of B. mori. As not all genes are interrupted, this case might be analogous to that of diptera. We have recovered two classes of insertions with unrelated sequence as judged by restriction enzyme analysis. Members of one class occur within the 28S gene in a comparable location to the predominant introns or insertions found in other organisms. Members of the second class interrupt the rDNA repeat at a point close to the 3' terminus of the 28S gene and may be interpreted in several different ways.

MATERIALS AND METHODS

Formation and screening of the cDNA library and the chorion gene-enriched genomic DNA sublibrary have been previously described (19-21). High molecular weight DNA was prepared by the liquid nitrogen-SDS-proteinase K method (22) from female silkmoths [strain 703 (23)] that had already laid their eggs. Southern hybridization of chromosomal or cloned DNA was performed as described (21). All hybridizations were conducted at 65°C in 0.6M salt [buffer S (0.15 M NaCl/0.03M Tris-HCl,pH8/1 mM EDTA) 4 times concentrated] with final washing of the nitrocellulose paper at 65°C in buffer S at 1/10 strength. After exposure of genomic Southern blots, X-ray films were scanned with an LKB-Soft Laser Scanning Densitometer (kindly provided by M. Gorovsky).

RESULTS

Isolation of rRNA specific cDNA clones

A cDNA library was constructed by using as starting material total RNA from the first five choriogenic follicles of each ovariole (19). This library contained DNA sequences corresponding to non-chorion genes, early chorion genes, as well as rRNA. Because rRNAs were by far the most abundant RNA species in the starting material, approximately 50% of the reverse transcripts corresponded to rRNA sequences, even though cDNA synthesis was primed with oligo(dT). Since the fraction of

reverse transcripts corresponding to rRNA were of limited complexity, while the remaining transcripts (50%) corresponded to a diverse population of mRNA species, rRNA clones were conveniently identified as the strongest hybridizing clones when screened with labelled cDNA derived from total cellular RNA. Direct demonstration that these clones contained rRNA sequences was obtained by their ability to hybridize exclusively to rRNA in a total cellular RNA blot (24; data not shown). The cDNA clones corresponding to rRNA inserts were divided into three groups of non-crosshybridizing sequences, typified by r101, r102 and r103.

Isolation of rDNA genomic clones

A sublibrary of 527 genomic clones, used for a "walk" in the chorion locus of *B. mori*, has been established by screening a genomic library with cDNA reverse transcribed from poly(A)$^+$ RNA representative of all chorion developmental stages (21). The genomic library was constructed in charon 4 phage after partial Eco R1 digestion of chromosomal DNA, therefore all Eco R1 sites found in the recombinant clones represent real genomic Eco R1 sites (25). The majority of the clones in this sublibrary (381 clones) are known to contain chorion genes of either "middle" or "late" developmental specificity (ref. 21 and T.H. Eickbush, unpublished data). The gene sequences on the remaining 146 clones should contain either additional uncharacterized chorion genes, non-chorion genes or rRNA genes. Assuming that at least some of these genomic clones would be of this latter class, r101 was used as probe to screen the genomic sublibrary. Surprisingly, 102 of the 146 uncharacterized clones hybridized, and thus represented a large pool of isolated rDNA repeats.

Eleven r101-positive genomic clones were selected at random for further analysis; three of these are identified in Figure 1 as B98, B108 and B113. Restriction mapping and Southern analysis, using r101, r102 and r103 as probes for various segments of the rRNA genes, revealed that nine of the clones, typified by B108, had identical 10.8 kb inserts. The restriction map of B108 was in complete agreement with the restriction map for the rDNA repeat reported by Manning et al (18). By reference to the latter map, the three classes of cDNA clones were identified as follows: the r101 class corresponds to segments

from the 5' half of 28S rRNA; the r102 class corresponds to segments from the 3' half of the 28S rRNA gene; and finally the r103 class corresponds to segments from the 18S gene. Clones B98 and B113 appeared to be partially overlapping with B108 and partially distinct. In particular, B113 did not contain sequences corresponding to the 18S gene probe (r103), while B98 did not contain the 3' half of the 28S gene (probe r102).

Presence of Insertions

The recovery of nine copies of B108 in a sample of eleven clones indicates that the rDNA of *B. mori* is predominately characterized by a 10.8 kb tandemly repetitive, essentially invariant unit as suggested previously (18). However, the existence of clones B98 and B113 suggests that this tandem unit is occasionally disrupted. To eliminate the possibility that these atypical rDNA clones were artifactually produced during construction or propagation of the recombinant phages, genomic blots of *B. mori* DNA, restricted to completion with various enzymes, were probed in the cDNA clone r101. If the rDNA repeat is invariant, those enzymes which cut only once within the repeat, such as Eco Rl and Hind III, should give rise to a single band on genomic Southerns equal in length to the B108 insert (neglecting the junction fragments at the end of the array, which should be less than 1% of the total if all rDNA repeats form a single array). As shown in the Eco Rl digest in Figure 2 (lane B), while the predominant genomic fragment which hybridizes to r101 is indeed 10.8 kb, corresponding to the standard repeats, three minor bands are also observed. These minor fragments are best explained (see Discussion) as rDNA repeats that contain non-ribosomal DNA inserts, with or without Eco Rl sites. In the case of Hind III digested genomic DNA, minor bands in addition to the 10.8 kb fragment are again observed (lane C). These minor bands have similar aggregate intensity but different size than the Eco Rl generated fragments. Different results were obtained with Bgl II, which cleaves three times within the rDNA repeat (Figure 1). Genomic DNA digested with Bgl II resulted exclusively in 5.1 and 0.5 kb fragments which hybridized to r101 (Figure 2, lane D). These fragments are identical to those predicted from the tandem

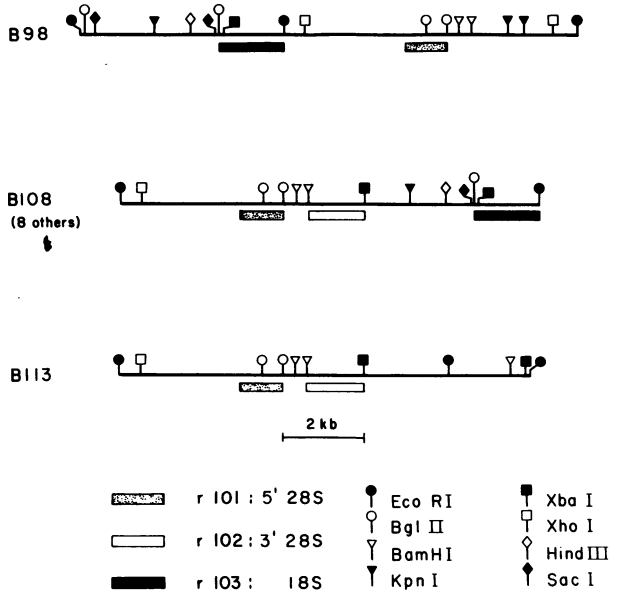

Figure 1. Restriction maps of the rDNA inserts of genomic clones B98, B108 and B113. The locations of the genomic segments homologous to the ribosomal cDNA clones r101, r102 and r103 were inferred by Southern hybridizations and are identified by boxes below the restriction maps. The location of r101 is known precisely, since the cDNA clone contains the leftmost Bgl II site of B108 as drawn.

duplication of clone B108, suggesting that no insertions into the rDNA repeat occur within these fragments.

Location of the insertions

A map of the rDNA repeat, indicating the approximate location of the insertions, is shown in Figure 3. The location of the 18S, 5.8S and 28S genes as determined by Manning et al (18) is also presented. Clone B113 is identical to the typical rDNA repeat between the Eco R1 site near the 3' end of the 18S gene and a position beyond the Xba I site near the 3' end of the 28S gene. An interruption in the rDNA repeat occurs between this Xba I site and the Kpn I site within the nontranscribed spacer. Exact positioning of the interrupting non-ribosomal DNA within the transcribed or non-transcribed portion of the repeat is not possible at this time. Only the left end of this inserted DNA

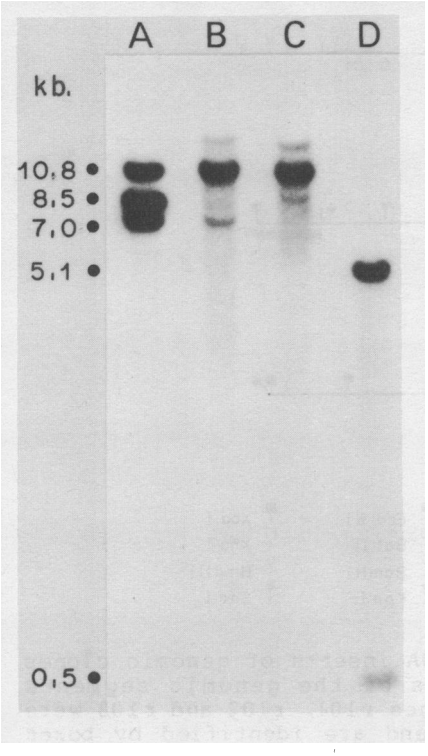

Figure 2. Genomic DNA blot analysis of the rDNA repeats of B. mori. (Lane A) Standards of approximately 0.1 µg each of DNA from clones B98, B108 and B113, digested with Eco Rl and yielding r101-hybridizable bands of 7.0, 10.8 and 8.5 kb, respectively (see Figures 1 and 3). (Lane B) 5 µg adult B. mori DNA digested with Eco Rl. (Lane C) as B but digested with Hind III. (Lane D) as B but digested with Bgl II. The DNAs were fractionated on a 0.7% agarose gel, blotted in parallel, and hybridized to nick-translated r101. The film used for this Figure was somewhat overexposed, so as to reveal the fragments corresponding to minor rDNA variant repeats.

has been cloned, because the DNA contains an Eco Rl site within approximately 3 kb from the insertion site. Genomic rDNA repeats containing this type of insert would be digested by EcoRl into an 8.5 kb fragment, thus accounting for one of the minor bands visible on genomic Southern blots probed with r101 (Lanes A and B of Figure 2). More recently, another clone with an insertion in the same general location was recovered (B. Robins and T. Eickbush, unpublished observations). Its insertion is apparently located between the Kpn I and Hind III sites of the non-transcribed spacer. In this case, the cloned Eco RI fragment is 10.2 kb, and thus would not be distinguishable from the predominant 10.8 kb band in genomic Southerns.

Clone B98 contains portions of two consecutive rDNA repeats, each of which apparently contains an interruption. The insert at the left end of the clone disrupts the rDNA repeat near the 3' end of the 28S gene in a manner similar to that seen in B113. In this case however, the right end of the non-ribosomal DNA insert

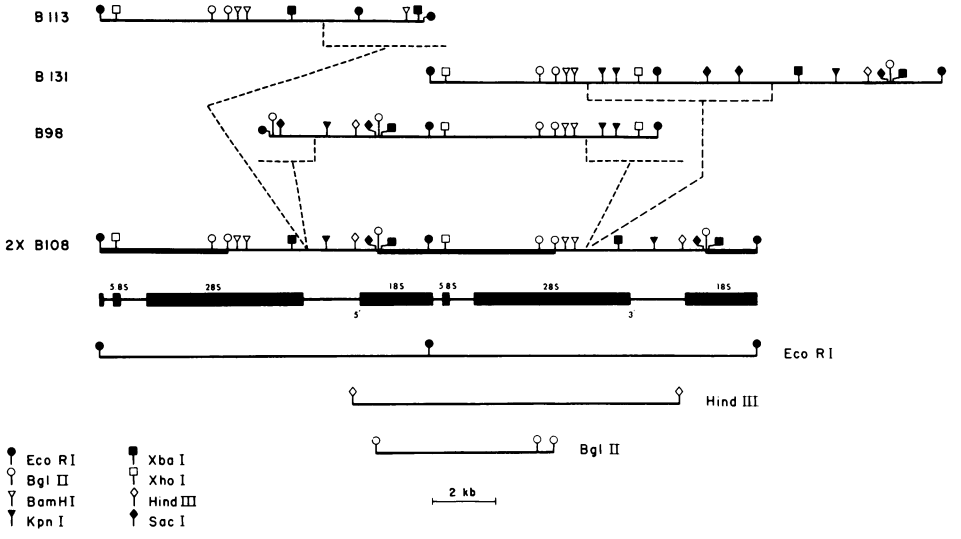

Figure 3. Restriction map of two consecutive rDNA repeats, indicating the location of insertions (dashed lines) within clones B113, B131 and B98. The locations of 18S, 5.8S and 28S genes as determined by Manning et al.(18) are drawn below the B108 map. Restriction fragments of the standard repeats which would be detected in genomic blots probed with r101 are shown at the bottom, as an aid in the analysis of Figure 2. The portion of the rDNA which shows no detectable inserts (according to Figure 2, lane D) is indicated with a thicker line in the restriction map of standard (B108) repeats.

is recovered. It is not known whether clones B113 and B98 represent the left and right ends, respectively, of the same class of rDNA inserts, or whether they represent different classes of inserts located in approximately the some position. The second rDNA insert present in B98 is located within the 28S gene, between the pair of Bam Hl sites and the Xba I site. This is the same region of the rDNA repeat where insertions are found in several other eukaryotic species (16). Exact localization of the insertion point will require DNA sequence analysis. This second rDNA insert accounts for the 7.0 kb Eco Rl fragment seen on genomic blots probed with r101 (Figure 2, lanes A and B).

Two additional clones with insertions in the same region were recovered recently (B. Robins and T. Eickbush, unpublished observations). One of these, B131, is represented in Figure 3. It is identical to the right half of B98 and extends further to

the right; thus it reveals that the total insertion is 4.9 kb and is flanked on either side by the two halves of the 28S gene. The second clone has a different insertion in the same general location, again flanked by 28S sequences on either side; in that case the relevant Eco RI fragment is 16 kb, and thus accounts for the top band seen in Figure 2, lane B.

In summary, the non-standard clones recovered to date represent two or three different insertions in one region, near the 3' end of the 28S gene, and two different insertions in another region, near the middle of the 28S gene. The DNA sequences of the recovered portions of these various insertions are not related, at least at the level of restriction enzyme analysis. By contrast, no detectable insertions occur within the 5.1 and 0.5 kb Bgl II fragments of the rDNA repeat (Figure 2, lane D), i.e. in the continuous segment of the repeat between the Bgl II site near the 5' end of the 18S gene and the pair of closely spaced Bgl II sites within the 28S gene (thick line in the B108 map, Figure 3).

To determine the percentage of rRNA cistrons in B. mori containing insertions, several autoradiographic exposures of genomic DNA blots similar to that shown in Figure 2 were scanned. With the Eco Rl digested genomic DNA, peaks corresponding to the 7.0, 8.5 and 15 kb fragments were pooled as representatives of those clones containing insertions, and compared to the 10.8 kb peak. By this determination approximately 12% of the rDNA cistrons contain inserts (approximately 7% for the B98/B131 type of insert). A slightly lower percentage of insertions (10%) was calculated using the two minor bands visible in the Hind III digested genomic DNA. These calculations represent the lower limit for the percentage of repeats with insertions, since they do not take into account the possibility of cistrons containing small inserts (< 0.5 kb) or inserts which coincidentally give rise to 10 - 11.5 kb fragments when digested with Eco Rl or Hind III and are thus hidden on the blot by the major repeat fragment. As mentioned above, one clone of the latter type has in fact been recovered.

DISCUSSION

The results presented in this paper suggest that at least 12% of the 240 rDNA cistrons of *B. mori* contain DNA insertions. These inserts fall into at least two classes by location and four by restriction map. One class of inserts interrupt the repeat near the middle of the 28S gene in a region similar to that seen in several other organisms. DNA sequencing studies of two types of rDNA inserts in *D. melanogaster* and one type of insert in *Tetrahymena pigmentosa*, have indicated that while these inserts occur within the same narrowly defined region of the 28S gene, their exact locations are distinctly different (26-28). Thus, sequencing studies are needed to determine if the *B. mori* insert position(s) correspond(s) precisely with that of any other species. Such studies are currently in progress. The second class of *B. mori* rDNA inserts lies close to the 3' end of the 28S gene. Dawid et al. (29) have also reported the isolation of a *D. melanogaster* rDNA repeat with an insertion near the 3' end of the 28S gene. While this type of insert is extremely rare in *Drosophila*, it represents one of the major insert types of *Bombyx*. Again, sequencing will be needed to ascertain whether these inserts are located at identical sites in *Drosophila* and *Bombyx*. Sequencing will also reveal whether any of these inserts are found within the transcribed portion of the 28S genes, or exclusively in the non-transcribed spacer, where one apparently is located.

Inserts of the first class, located in the middle of the 28S gene, have been shown to be flanked, within a single clone, by complementary portions of the rDNA repeat. Thus, a majority of all apparent inserts are *bona fide*. Interpretation of the second class is more ambiguous, pending partial sequence analysis and recovery of individual clones encompassing an entire insert plus flanking sequences. In addition to inserts in the 3' end of 28S genes or in the non-transcribed spacer, this second class might include the ends of the rDNA repeat arrays, and even isolated rDNA repeats. We note, however, that a single nucleolar organizer exists in *B. mori* (30,31). Furthermore, the 8.5 Eco RI band in genomic Southerns, which is of the same size as the B113 clone insert, corresponds to approximately 4 copies per haploid

genome. It is unlikely that this band corresponds to isolated rDNA repeats or ends of repeat arrays, because such repeats would have to be embedded in similar flanking sequences to yield multicopy genomic DNA bands. Finally, as already mentioned, there is a precedent for such an insert location in Drosophila (29).

One possible explanation for the failure of Manning et al. (18) to detect variant rDNA repeats is their remarkable variance in different inbred strains of B. mori (unpublished observations). For example, nearly 20% of the rDNA cistrons of strain C108 (32) contain insertions, belonging to at least 5 different classes. It is possible that the strain used by Manning et al. had considerably fewer variant repeats, which thus were not easily detected in genomic DNA blots.

ACKNOWLEDGEMENTS

We thank Berklee Robins for permission to quote unpublished results. This work was supported by NSF, ACS and March of Dimes grants to F.C.K., and by an NIH grant to T.H. Eickbush.

Addresses: [3]Department of Biology, University of Rochester, Rochester, NY 14627, USA, and [4]Institute of Molecular Biology and Biotechnology and Department of Biology, University of Crete, Heraclio, Crete, Greece

REFERENCES

1. Long, E.O. and Dawid, I.B. (1980) Annu. Rev. Biochem. 49, 727-764.
2. Cambell, G.R., Littau, V.C., Melera, P.W., Allfrey, V.G., and Johnson, E.M. (1979) Nucleic Acids Res. 6, 1433-1447.
3. Gubler, U., Wyler, T., and Braun, R. (1979) FEBS Lett. 100, 347-350.
4. Wild, M.A. and Gall, J.G. (1979) Cell 16, 565-573.
5. Din, N., Engberg, J., Kaffenberger, W., and Eckert, W. (1979) Cell 18, 525-532.
6. Bos, J.L., Heyting, C., Borst, P., Arnberg, A.C., and Van Bruggen,E.F.J. (1978) Nature 275, 336-338.
7. Rochaix, J.D. and Malnoe, P. (1978) Cell 15, 661-670.
8. Wellauer, P.K. and Dawid, I.B. (1977) Cell 10, 193-212.
9. Glover, D.M. and Hogness, D.S. (1977) Cell 10, 167-176.
10. Barnett, T. and Rae, P.M.M. (1979) Cell 16, 763-775.
11. Beckingham, K. and White, R (1980) J.Mol. Biol. 127, 349-373.
12. Renkawitz-Pohl, R., Glatzer, K.H., and Kunz, W. (1981) J. Mol. Biol. 148, 95-101.
13. Renkawitz-Pohl, R., Matsumoto, L., and Gerbi, S.A. (1981) Nucleic Acids Res. 9, 3747-3778.

14. French, C.K., Fouts, D.L., and Manning, J.E. (1981) Nucleic Acids Res. **9**, 2563-2576.
15. Long, E.O. and Dawid, I.B. (1979) Cell **18**, 1185-1196.
16. Beckingham, K (1982) in rDNA, Part A. The Cell Nucleus, 10, H. Busch and L. Rothblum, eds. (New York, Academic Press), pp.205-269.
17. Gage, L.P. (1974) J. Mol. Biol. **86**, 97-108.
18. Manning, R.F., Samols, D.R. and Gage, L.P. (1978) Gene 4, 253-166.
19. Lecanidou, R., Eickbush, T.H., Rodakis, G.C., and Kafatos, F.C. (1983) Proc. Natl. Acad. Sci. USA **80**, 1955-1959.
20. Eickbush, T.H., Jones, C.W., and Kafatos, F.C.(1981) in Developmental Biology Using Purified Genes. ICN-UCLA Symposia on Molecular and Cellular Biology, **23**, D. Brown and C.F. Fox, eds (New York, Academic Press), pp. 135-153.
21. Eickbush, T.H. and Kafatos, F.C. (1982) Cell **29**, 633-643.
22. Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303-2308.
23. Iatrou, K., Tsitilou, S.G., Goldsmith, M.R., and Kafatos, F.C. (1980) Cell **20**, 659-669.
24. Alwine, J.C., Kemp, D.J., and Stark G.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5350-5354.
25. Maniatis, T., Hardison, R.C., Lacey, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K., and Efstratiadis, A. (1978) Cell **15**, 687-701.
26. Wild, M.A. and Sommer, M. (1980) Nature **283**, 693-694.
27. Dawid, I.B. and Rebbert, M.L. (1981 Nucleic Acids Res. 9, 5011-5021.
28. Roiha, H., Miller, J. R., Woods, L.C. and Glover, D.M. (1981) Nature **290**, 749-753.
29. Dawid, I.B., Wellauer, P.K., and Long, E.O. (1978) J. Mol. Biol. **126**, 749-768.
30. Rasmussen, S.W. (1976) Chromosoma **54**, 245-293.
31. Traut, W. (1976) Chromosoma **58**, 275-284.
32. Goldsmith, M.R. and Basehoar, G. (1978) Genetics **90**, 291-310.