Structure and function of an AT-rich, interspersed repetitive sequence from Chironomus thummi: solenoidal DNA, 142 bp palindrome-frame and homologies with the sequence for site-specific recombination of bacterial transposons

Norbert Israelewski

Ruhr-Universität Bochum, MA 5, Lehrstuhl für Genetik, D-4630 Bochum, FRG

Received 19 July 1983; Revised and Accepted 3 October 1983

ABSTRACT

Chironomus thummi thummi contains a repetitive AT-rich 118 bp sequence mainly in the centromere regions and elsewhere in the genome (1). A large cluster of repeats is regularly present in the non-transcribed spacer of rDNA. Dimers and multimers of the repeat migrate slower in small pore gels than would be expected from their size. The results indicate a solenoidal structure with a coil girth of appr. 350 bp. This structure is most probably due to a highly periodic positioning of di-nucleotides of the type purine - purine or pyrimidine-pyrimidine with distances of appr. 10 bases. In a cluster of 118 bp repeats, regions of dyad-symmetry are positioned such that a 142 - 2 bp palindrome-frame is generated. Evidence is presented favouring the assumption that the repeat functions primarily in sister chromatid exchange.

INTRODUCTION

Among Chironomides the two subspecies Chironomus thummi thummi and Chironomus thummi piger have received special interest since they represent a system which allows the first steps in species separation to be studied. This separation process is accompanied by a geometric increase of DNA in polytene chromosome bands mainly at the centromeres and neighbouring regions of the Chironomus thummi thummi genome (2, 3). The local increase of DNA is in some way coupled with an increase in the amount of repetitive DNA sequences (4, 5) which is also visible in an increase of the amount of C - banding DNA (6). A comparable situation is found in Drosophila sibling species where it has been emphasized that apparent differences in the amount of highly repetitive DNA sequences accompanies species separation (7, 8). The variation in the highly repetitive portion of the genome might also include regions carrying information for proteins (9).

In the two Ch. thummi subspecies a centromeric cluster of an AT-rich repetitive 118 bp sequence has been characterized (1). This cluster has been magnified and dislocated in Ch. th. thummi and is also found at those sites where the DNA content in bands has increased (1). The fact that this sequence
Nucleic Acids Research

has also entered the non-transcribed spacer of rDNA led us assume that the repeti-
tive sequence may function in promoting sister chromatid exchange (SCE) (10). A high frequency of SCE would provide the redundant gene cluster with a greater evolutionary flexibility (11). Evidence has been accumulated favour-
ing this hypothesis (12, 13 and this report).

MATERIAL AND METHODS

1. DNA

Cloned rDNA of Ch. th. thummi pCtt 1505 and pCtt 1507 was kindly provided by E. R. Schmidt (17). The plasmid DNA was further purified from contaminating RNA by ethanol precipitation in the presence of 2.5 M ammonium acetate. If desired, the rDNA insert was separated from the vector pBR 328 by EcoRI restric-
tion of the clone followed by preparative gel electrophoresis and ele-
troelution of the DNA.

2. DNA Restriction and Gel Electrophoresis

Enzyme incubations were performed as described by the producer (Boehrin-
ger). Restriction fragments were separated on 3 mm horizontal slab gels in a tris - phosphate - EDTA buffer system (14). Agarose gels were prepared accord-
ing to (15, 16). Polyacrylamide gels ( 1 : 20 ratio N,N'methylenbisacryl-
amide : acrylamide) were reinforced with 0.5% agarose. OX174 - HaeIII and K- HindIII fragments were used as size markers and the gels analysed by ethi-
dium bromide staining. Negative prints are shown.

RESULTS

The restriction map of the cloned rDNA of Ch. th. thummi pCtt 1507 is shown in figure 1. This clone contains appr. 22 x 118 bp repeats (Clal re-
peats) which has been sequenced (17). The map is identical with that of pCtt 1505 except that clone pCtt 1507 exhibits stability in the number of Clal
repeats after replication in E. coli unlike pCtt 1505 which is highly un-
stable in the number of Clal repeats (17). Sequence data have shown that the stable clone 1507 contains a base substitution at the right end of the Clal
repeat cluster (18) and several uncharacterized mutations in the vector pBR
328 including the spread of Clal repeats into the vector (Israelewski, unpubl). It was therefore decided to separate the rDNA insert from the vector by EcoRI restriction and preparative gel electrophoresis.

Measurements of the length of the EcoRI fragment based on its mobility on gel electrophoresis gives rise to different values which are apparently dependent on the pore size of the gel. Higher percentage gels slow down dis-
Fig. 1 Restriction map of a cloned Ch. th. thummi rDNA cistron pCtt 1507 (according to 10. 17). Of the Hpall sites only the two in the NTS are shown. The asterix indicates a ClaI site that is present in most of the genomic rDNA cistrons, but is absent in cloned rDNA.

proportionately the migration of the fragment. In 0.5%, 1.0%, 1.5% and 2.0% agarose gels lengths of 11.0 kb, 12.9 kb, 15.5 kb and 40 kb were measured respectively. When HaeIII restricted DNA is electrophoresed it is only the NTS carrying fragment which migrates anomalously. The same is true for HpaII restriction fragments. (fig. 2A). After ClaI digestion it is seen that the two flanking segments of the ClaI repeat cluster migrate correctly in 2.0% agarose gels. Comparisons of the gel dependent shift in the sizes of the EcoRI, HaeIII and HpaII fragments indicate that the shift is more pronounced if the ClaI repeat cluster is flanked by large DNA segments which alone do not show the shift (fig. 2B). In the case of the HpaII fragment there additionally appears a faint band without shift (arrows, fig 2A, clearly seen only in 1.5% and 2.0% agarose gels). Since no HpaII fragment greater than 2.6 kb should exist besides the NTS-HpaII-fragment (3.7 kb) it is possible that in few plasmids one HpaII site is modified (at the left of 28S) resulting in the 3.7 kb band.

For analysis of the gel-type dependent shift in the sizes of monomers and multimers of the ClaI repeat the clone pCtt 1507 was digested incompletely with ClaI and the DNA was run on polyacrylamide (PAA) gels. The shift is measured as percentage deviation from the real value. Three typical curves are shown in figure 3A. In all three PAA gels (2.0% - 2.5% - 3.0%) the monomer size was measured to be 116 bp corresponding to a deviation of -2% relative to the 118 bp fragment of ØX174-HaeIII. With 2.0% gels the shift of dimers and multimers increases up to +20% for six ClaI repeats. Further increasing of the number of the repeats does not contribute to a further increase of the shift. With 2.5% PAA gels the curve shows systematic steps resulting in kinks of the curve at intervals corresponding appr. to multiples of three ClaI repeats. With 3.0% PAA gels the shift increases dramatically up to nine repeats; further multimers cannot be measured in this type of gel.

In each case care was taken to avoid partial denaturation of the DNA which alters its mobility on electrophoresis. If the gel heated up due to a
high current, the curves shown above are not exactly reproducible. However, partial denaturing of the Clal repeat DNA under controlled conditions (2M urea in the sample application buffer) leads to a systematic enlargement of the steps in the curve for 2.5% PAA gels, especially at higher numbers of Clal repeats (fig. 3B).

It has been noticed that clone pCtt 1505 is not stable during replication in E. coli leading to a decrease in the number of Clal repeats (17). The presented data of a secondary structure of the Clal repeat and the possible interference of a recombination system of E. coli with the Clal repeat sequence (17 and discussion) prompted us to re-evaluate the elimination process of Clal repeats in E. coli. Using a single colony isolate for the DNA preparation, the restriction with HaeIII results in a ladder of NTS fragments with size intervals of 120 bp (fig. 4A). It is interesting that certain fragment sizes are preferred during the elimination process in E. coli. In the lower molecular weight region it is seen that these fragments have distances of 3 x 120 bp or multiples of that (arrows). If E. coli is cultivated successively four times overnight, three prominent HaeIII fragments accumulate which have also size intervals of 360 bp (3 x Clal repeat, fig. 4B). Thus, it is suggested that recombination in the cloned rDNA occurs at a defined site in the Clal sequences modulated by the structural feature of the DNA.

DISCUSSION

Structure and function are two aspects of living matter (19). Considering the results suggesting a defined secondary structure of the Clal repeat DNA of Ch. th. thummi it may be anticipated that one can also find a function. Thus, sequence data (17) were analysed.

I. 10 bases periodicity and bent helical DNA structure

Trifonov (20) has predicted that the DNA axis is curved if some dinucleotides of the type purine-purine or pyrimidine-pyrimidine have the tendency to be repeated with a period of about 10 bases. In a long DNA fragment this would lead to a solenoidal DNA structure which is stable in a DNA molecule free of protein. In the Clal repeat sequence an almost perfect 10 bases

Fig. 2 A) Agarose gel electrophoresis of the rDNA insert of pCtt 1507 after restriction with the enzymes indicated at the top of each lane. The NTS containing fragments (•) migrate anomalously in 1.0% - 2.0% agarose gels. On elimination of the Clal repeat DNA by digestion with Clal the two flanking fragments (•) of the repeat cluster migrate correctly in 2.0% agarose gels. Arrow: see text. B) Plot of the measured size of the NTS containing fragments as a function of the percentage of agarose.
Fig. 3 Gel-type dependent shift (%) in the apparent sizes of monomers and multimers of the Clal repeat on electrophoresis in 2.0% - 3.0% polyacrylamide gels. The shift is plotted against the number of Clal repeats. The fragment at 0.34 kb (*) is derived from 28S rDNA and it does not show the shift. Sample application buffer contained A) 1 M urea, B) 2 M urea.

periodicity is visible (fig. 5). In accordance with Trifonov's findings is the fact that purine-purine dinucleotides have a maximum when the pyrimidine-pyrimidine ones have a minimum. Marini et al. (20) have shown that the se-
sequence dependent curving of DNA of the type above results in a too slow migration of the DNA structure in small pore gels. Here it is demonstrated that flanking 'linear' DNA segments enhance the anomalous migration. Furthermore, on investigation of the migration of monomer and multimers of the Clai repeat a stepwise altering of the mobility of the DNA is observed every 350 bp (appr. 3 x repeat length). It is likely that this periodicity reflects the coil girth of the DNA under consideration, since it is this parameter of a spiral which is coupled with periodic repetitions of a certain DNA length.

The observed periodic alteration in the mobility of the Clai repeat DNA structure may be explained statistically in that each coil contributes to align the DNA structure at a pore of the gel. If so, the actual pore size of a gel determines the degree of alignment which is necessary for the DNA
Fig. 5 Distribution of the dinucleotides of the type A) purine-pyrimidine, B) purine-purine, C) pyrimidine-pyrimidine. Frequencies of the occurrence of the dinucleotides were recorded within distances of three bases along the 118 bp sequence.

structure to snake through the pore. Once aligned, further coils passing the pore do not significantly slow down the migration of the DNA structure (fig. 3)

II. 142 bp distances between regions of dyad-symmetry in the clustered repeat

It has been suggested that the sequence-dependent curving of the DNA axis facilitates its folding in chromatin (22). In this context it seems noteworthy that other potential structural elements, regions of dyad-symmetry, are obviously coupled with a systematic positioning message present in the Clal repeat DNA. It is shown in figure 6 that in a cluster of seven Clal repeats palindromes are positioned such that a 142 ± 2 bp frame is generated, this

Fig. 6 Regions of dyad-symmetry over at least 5 bp without mismatch positioned in a cluster of seven Clal repeats (schematically delimited by bars). Numbering of the bases refers to Schmidt et al. (17). Only those palindromes are shown which are positioned in the 142 ± 2 bp frame (circled bases).
Fig. 7 Regions of dyad-symmetry over two 234 bp Alul repeats of Drosophila melanogaster (schematically delimited by bars). Numbering of the bases refers to Miller et al. (24). Only those palindromes are shown which are positioned in the 142 bp frame (circled bases).

being in excellent agreement with the core DNA length in a nucleosome. The DNA looped out by the palindromes is not constant in length but has, in most cases, lengths of appr. 30 bp. It is evident that such a frame cannot evolve by chance and we have no explanation other than the suggestion that nucleosomes select their binding sites by utilizing these structural features. The positions of the 142 bp-frame palindromes in the solenoid appears to be not entirely random: appr. each half turn of a coil would be the potential site for palindromic refolding. The palindromes shown in figure 6 represent one third of the number of palindromic sequences present in the ClaI repeat. A detailed description of all palindromic sequences together with the statistical analysis is in preparation. Briefly, a systematic arrangement is found for all 15 palindromes utilising the basic frame of 142 ± 2 bp.

In the african green monkey 172 bp repeat which is known to be phased with nucleosomes (23, 49) the perfect palindromic sequence GATATTT - 31bp - AAATATC is capable of looping out the presumed linker DNA region thus generating a 141 bp frame in the clustered repeat. Recently the positions of the phased nucleosomes on the 172 bp repeat have been determined (49). The nucleosome core boundaries of the most frequent frame F fit the above mentioned palindrome with an accuracy of 3 bp.

Another example of systematically arranged palindromes is present in the 234 bp repeat of the rDNA of Drosophila melanogaster. These repeats serve as 'loading sites' for RNA polymerase I and they can augment the transcription of the rDNA unit (24, 25, 26). This frame, however, is somewhat differently organized when compared to the one in Ch. th. thummi (fig. 7).
In conclusion, these examples lead one to believe that palindromes in the 142 bp frame are important informational structures for assembly and/or function of repeated DNA sequences. In addition, the palindrome frame would not allow that a random variation in the repeat size can be accepted.

Fig. 8 Sequence homologies between the 118 bp Clal repeat (numbering of the bases refers to Schmidt et al. (17) and the sequence for site-specific recombination (resolution) of the bacterial transposon TN3 (Grindley et al.,27). Underlined are the three binding sites for the bacterial resolvase enzyme. The actual site of recombination in TN3 is indicated by the arrow. Homology is 64%.

In conclusion, these examples lead one to believe that palindromes in the 142 bp frame are important informational structures for assembly and/or function of repeated DNA sequences. In addition, the palindrome frame would not allow that a random variation in the repeat size can be accepted.

III. Homologies with the sequence for site-specific recombination of bacterial transposons of the type TN3

It appears highly surprising that the Clal sequence has 64% homology with the bacterial sequence which is known to mediate site-specific recombination of transposons Yd and TN3 during DNA replication in E. coli (fig. 8). But with this fact in mind four phenomena coupled with the presence of Clal repeats of Ch. th. thummi become explainable: 1. The Clal repeat is found at sites where a geometric increase of DNA segments have occurred (1, 3), 2. An extensive variation in the number of Clal repeats is found in the NTS regions of rDNA cistrons (10). 3. It is difficult to clone the rDNA of Ch. th. thummi in E. coli. One of the two clones obtained is highly unstable in the number of Clal repeats during replication in E. coli (17 and results). The second clone is stable, but this is coupled with several mutations (see results). 4. In certain crosses of Ch. th. thummi x Ch. th. piger a high incidence of chromosomal mutations is observed at sites with high number of Clal repeats (28).

Common to all these phenomena is that some form of exchange is necessary. Phenomena 1 - 3 involve unequal exchange and phenomenon 4 illegitimate exchange. In conclusion, there is growing evidence that the Clal repeat together with its structure serves a role in the generation of sister chromatid ex-
change (SCE) which might be unequal. It is known that SCEs are produced during DNA replication (29,30,31). Unequal SCE is especially favourable for redundant genes in order to ensure the horizontal evolution within the gene cluster (11, 32). Additionally, the ClaI repeat DNA and possibly other related AT-rich sequences can punctuate genomic DNA so that the exchange event is necessarily confined to the repeats. This corresponds to the conception for meiotic exchange that chromosomal structures have arisen in a functional connection with exchange events in lampbrush chromosomes. Thereby, the gene regions are protected from disintegration (34,35,36). As a matter of fact, punctuation of the genomic DNA by AT-rich sequences is characteristic of eukaryotes (37,38).

Finally, it should be emphasised that the present investigation adds evidence that repetitive DNA sequences can be involved in recombination systems (39 - 44). In plant species a family of conserved middle repetitive DNA sequences of modal lengths of 1 - 2 kb plays an important role in meiotic recombination (45,46,47). A similar sized DNA sequence (1.1 kb) is apparently involved in the meiotic exchange process of the Diptera Phryne cincta (Israelewski, in prep.). Furthermore, it appears that at least some of the recombination mechanisms have been conserved essentially before pro- and eukaryotes diverged. This can also be deduced from the fact that a repetitive DNA sequence from the kangaroo rat Dipodomys ordi is recognized by the recombination system RecE of E. coli (48). In this case the authors emphasise that this repetitive sequence may provide the animal with genomic plasticity via mechanisms of DNA exchange. In fact, the kangaroo rat adapt rapidly to low levels of selection (48). This is what is observed also with the world-wide distributed Ch. th. thummi (high copy number of ClaI repeats) when compared with its endemic living relative Ch. th. piger (low copy number). By the same argument, apparent alterations in the copy number of ClaI repeats in thummi can account for a first step in species separation.

ACKNOWLEDGEMENTS

I thank Prof. Dr. H.-G. Keyl for critical reading of the manuscript and Dr. E. R. Schmidt who provided me with the cloned rDNA of Ch. th. thummi. The help and advice of Dr. E. A. Godwin in preparing of the manuscript is greatly acknowledged. This research was supported by the Deutsche Forschungsgemeinschaft, grant Ke 41, 15.

*Dedicated to Prof. Dr. B. E. Wolf in honour of his 75th birthday on September 27, 1983

REFERENCES

41. John, B. (1973) Chromosoma 44, 123-146
44. Yamamoto, M. (1979) Genetics 93, 437-448

6996