The EMBO Journal Vol.2 No.7 pp.1177 – 1183, 1983

The nucleotide sequence of an unusual non-transcribed spacer and its ancestor in the rDNA in *Chironomus thummi*

E.R.Schmidt* and E.A.Godwin

Lehrstuhl für Genetik, Ruhr-Universität, D-4630 Bochum 1, FRG

Communicated by W.Beermann
Received on 29 March 1983; revised on 18 April 1983

The nucleotide sequence of the non-transcribed spacer (NTS) in the ribosomal DNA (rDNA) of *Chironomus thummi thummi* and *Chironomus thummi piger*, including major parts of the external transcribed spacer, is described. The NTS of the two subspecies are very different in length, (*thummi*, 7 kb, *piger*, 2 kb); this is due to the insertion into the NTS of *C. th. thummi* of a large cluster of highly repetitive DNA sequences which are not present in the NTS of *C. th. piger*. The repetitive sequences, called Cla elements, are present in high copy number elsewhere in the genome of *C. th. thummi* and, in lower copy number, in the genome of *C. th. piger* in which they are mainly in the centromeric regions. Sequencing of the NTS of *thummi* and *piger* yielded information on the junctions between the Cla element cluster and the original NTS sequence, as well as on the sequence of the integration site before the transposition has occurred. The integration site is characterized by a dA cluster at the one end and a dT cluster at the other.

Key words: Chironomus/NTS/rDNA/transposition/Z-DNA

Introduction

The repetitive genes for the ribosomal 18S, 5.8S and 28S RNAs are organized as repeat units containing a frequently transcribed part, which includes the coding regions for 18S, 5.8S and 28S RNAs, and a non-transcribed, or less frequently transcribed part, usually called non-transcribed spacer (NTS). While the coding regions seem to be evolutionarily rather conservative, the NTS is much more variable and differs significantly between closely related species (Cortadas and Pavon, 1982; Long and Dawid, 1980; Bosely *et al.*, 1979; Rae *et al.*, 1981). Even among different strains of the same species or different individuals of one strain there can be considerable length heterogeneity (Arnheim and Kuehn, 1979; Coen *et al.*, 1982; Kunz *et al.*, 1981; Long and Dawid, 1980; Rae *et al.*, 1981), which is due to the presence of short repetitive elements in variable numbers. These repetitive elements with a basic repeat length ranging from 61 (*Xenopus*, Boseley *et al.*, 1979) to 350 bp (*Calliphora*, Schäfer *et al.*, 1981) seem to be confined to the NTS of rDNA not being present elsewhere in the genome (Rae *et al.*, 1981). However, an exception to this rule are sequences in the NTS of the mouse and probably other mammals (Arnheim *et al.*, 1980), where spacer sequences are also found elsewhere in the chromosomes. In addition, we have recently found a highly repetitive DNA sequence in the DNA of *C. thummi*, which is present in the NTS of rDNA as well as in many other chromosomal locations (Schmidt *et al.*, 1982; Schmidt, 1981). This sequence is characterized by a recognition site for the restriction endonuclease *Cla*I, a basic repeat length of 117 ± 2 bp and by its predominant location in the centromeric regions of the *C. th. thummi* chromosomes (Schmidt, 1981; Schaefer and Schmidt, 1981).

Comparison of the two closely related sibling species or subspecies of *C. thummi, C. th. thummi* and *C. th. piger* revealed that the concentration of this repetitive sequence family, called the Cla element, in the two genomes is related to the genome size: the subspecies with the larger genome (*C. th. thummi,* Keyl, 1965) contains ~6 times more Cla elements than the subspecies with the smaller genome *(C. th. piger)* (Schaefer and Schmidt, 1981). In addition to this difference in the concentration of the Cla elements, a distinct difference in their distribution on the chromosomes is found by *in situ* hybridization. While the Cla elements in the *C. th. piger* chromosomes are confined to the centromeric regions, they are found in the *C. th. thummi* chromosomes outside the centromeres (Schmidt, 1981).

One example of the different location of the Cla elements in the *C. th. thummi* and *C. th. piger* chromosomes is the organization of the NTS of the rDNA. While the NTS of *C. th. thummi* rDNA contains many copies (variable numbers of up to >100) of the Cla element (Schmidt *et al.*, 1982), the NTS of *C. th. piger* rDNA is devoid of this repetitive element. Since it is known from cytological data that the *C. th. piger* karyotype is phylogenetically the older one (Keyl, 1962), one can assume that the NTS of *C. th. piger*, without the repetitive Cla elements, is similar to the ancestor of the NTS of *C. th. thummi*, and that the NTS of *C. th. thummi* developed by the insertion of the repetitive Cla element cluster into such a *piger*-like NTS. The availability of both the ancient and the modern type of NTS offers the opportunity to investigate in detail the structure of the integration site before *(C. th. piger)* and after *(C. th. thummi)* the insertion of the repetitive elements. Furthermore, one can obtain information on the junctions between the highly repetitive Cla element cluster and the neighbouring DNA. Therefore, we have sequenced the non-transcribed and the adjacent external transcribed spacer sequences from both subspecies, using ribosomal repeat units cloned previously in *Escherichia coli*.

Results

Restriction map and sequencing strategy

The detailed restriction maps of the ribosomal repeat units of *C. th. thummi* and *C. th. piger* are shown in Figure 1. The interesting regions containing the entire or nearly entire NTS-ETS regions are located between two *Hae*III sites, one located within the 28S coding region, close to the 3′ end of the 28S gene, the other located within the vector DNA, close to the *Eco*RI site, which has been used as the integration site for the ribosomal repeat units. This *Hae*III fragment contains a small region of the 28S gene, sufficient to hybridize with the 28S RNA (Israelewski and Schmidt, 1982), the entire NTS and probably the major part of the ETS (Schmidt *et al.*, 1982). Therefore, we decided to sequence this *Hae*III fragment in total.

The detailed restrictions maps shown in Figure 1 reveal a number of differences between the NTS of *C. th. piger* (Figure 1A) and *C. th. thummi* (Figure 1B). The most con-

*To whom reprint requests should be sent.

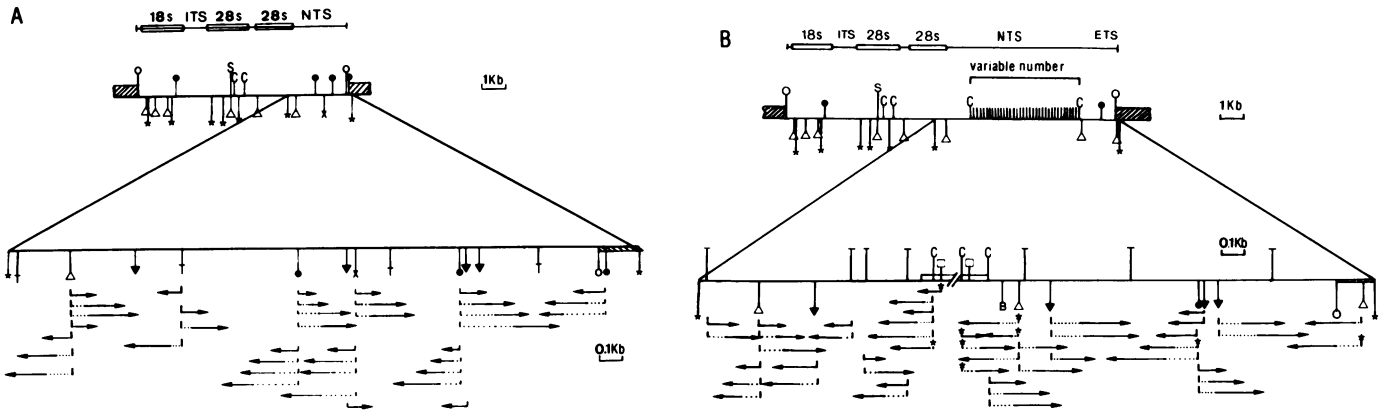

Fig. 1. Detailed restriction maps and sequencing strategy for the NTS of *C. th. piger* **(A)** and *C. th. thummi* **(B)**. The arrows summarize the sequencing gels that were read. The base of each arrow represents the labelled 3′ end and the arrowheads point towards the 5′ end. The dashed parts represent those sequences which have migrated out of the gel and the uninterrupted lines show the parts which were read unambiguously. A star at the base of the arrow indicates that these sequences were obtained from the 'second' rDNA clone pCtt 1507. The sequence in **B** is not contiguous within the repetitive Cla element cluster (open boxes), because of its enormous length (~5 kb in pCtt 1505 and ~3 kb in pCtt 1507). Hatched boxes = vector ○ *Eco*RI, ↓ *Hae*III, ● *Hind*III, ⚲ *Cla*I, ⚲ *Taq*I, ▮ *Bgl*II, ▮ *Pvu*II, ▼ *Hinf*I, △ *Hpa*II, ⚲ *Sau*3AI, ★ *Xba*I.

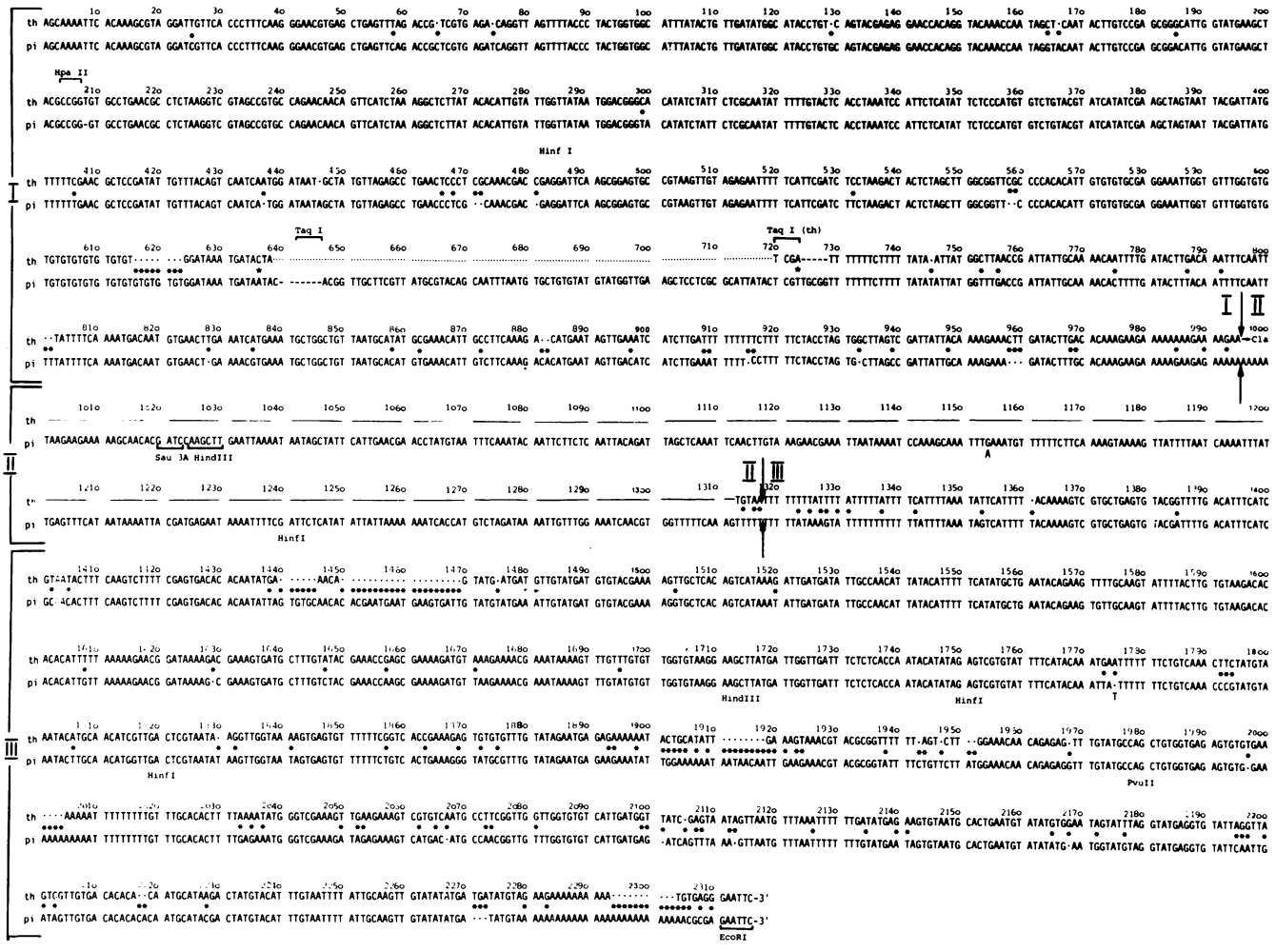

Fig. 2. Nucleotide sequence of the NTS of *C. th. thummi (th)* and *C. th. piger (pi)*. The nucleotide sequence is orientated in a 5′−3′ direction with the end of the 28S gene at the 5′ end and the start of the 18S gene at the 3′ end. Base no. 1 is located ~80 bases downstream from the *Hae*III site. The sequence of bases 1−50 possibly contains the end of the 28S gene (Israelewski and Schmidt, 1982). Restriction sites are indicated only when they are relevant to the sequencing strategy. The roman numerals refer to the sections I−III made in the NTS in Figure 4. In section I, a small fragment of ~80 bases was not sequenced (bases 638−719) in *C. th. thummi*. In section II (bases 995−1317), only the NTS of *C. th. piger* is shown; the corresponding region of *C. th. thummi* (——) containing the repetitive Cla elements and the Cla-adjacent DNA is shown in Figure 3. • = Deletions/insertions, ★ = divergencies between *C. th. thummi* and *C. th. piger*.

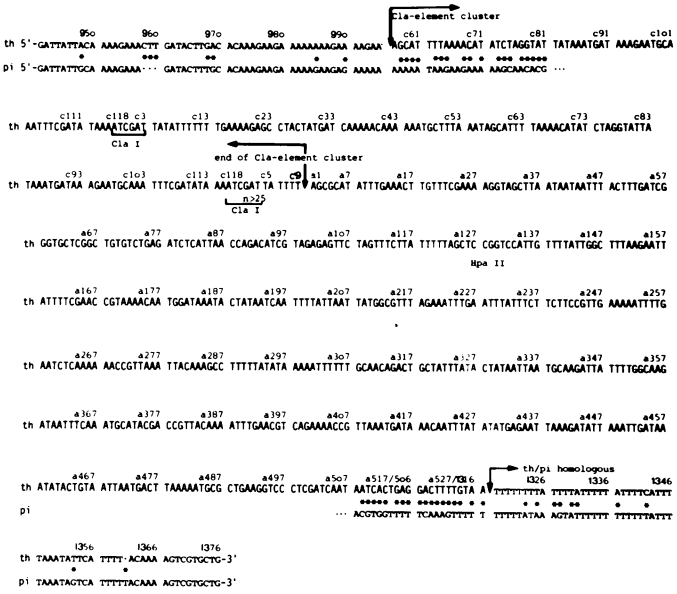

Fig. 3. Nucleotide sequence of region II (according to Figure 4) of the NTS of *C. th. thummi* including the repetitive Cla element cluster and the Cla-adjacent DNA. The sequence numbering of Cla elements starts with c57. Base c1 of the Cla element was defined to be G of the *Cla*I restriction site, that is characteristic for the Cla elements. Only the Cla elements that were actually sequenced are shown, although >25 Cla elements are present in this position and therefore the sequence is not contiguous within the Cla element cluster. The two DNA sequences were fitted into a continuous sequence by joining the left-most Cla element to the right-most Cla element of the Cla cluster. Downstream from the Cla elements the nucleotide sequence of the Cla-adjacent DNA (bases a1 – a528) is shown. The homology region in *C. th. thummi* and *C. th. piger* starts with base 1318 according to the numbering scheme in Figure 2. Small characters = homologous in *C. th. thummi* and *C. th. piger*, * = divergent in *C. th. thummi* and *C. th. piger*.

spicuous difference is the large cluster of *Cla*I sites present in the NTS of *C. th. thummi* and missing in the NTS of *C. th. piger*. The regularly spaced *Cla*I sites are due to the presence of a tandemly repeated DNA sequence, the Cla element, which has a basic repeat length of 117 bp. The number of Cla elements present in the NTS of *C. th. thummi* is variable, in the genomic DNA as well as in the cloned rDNA. Upstream from the Cla element cluster (towards the 3′ end of the 28S gene) the restriction maps of *C. th. thummi* and *C. th. piger* are identical. Downstream from the Cla element cluster (towards the 5′ end of the 18S gene) there is a region of ~300 bp in *C. th. piger* and ~500 bp in *C. th. thummi* displaying different restriction sites. The rest, containing probably mainly ETS, shows an identical restriction pattern for both subspecies.

Because of the great length of the repetitive Cla element cluster it was not possible to establish a contiguous sequence in this region. Two Cla elements located at the 3′ end of the cluster were sequenced using fragments labelled at the *Hpa*II site close to the Cla element cluster. However, at the 5′ end of the Cla element cluster, this strategy was not applicable, because an appropriate restriction site was not available. Nevertheless, part of the left-most Cla element was sequenced by exploiting the *Cla*I and *Sau*3AI restriction sites present in the repetitive Cla element. By this strategy we obtained the nucleotide sequence information about the boundary regions of the repetitive Cla element cluster. The rest of the sequencing strategy can be seen in Figure 1.

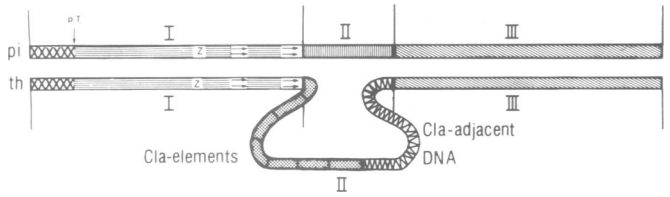

Fig. 4. Schematic representation of corresponding regions in the NTS of *C. th. piger (pi)* and *C. th. thummi (th)*: region I ranges from the end of the 28S gene to the site where the Cla elements are integrated into the NTS of *C. th. thummi* (XX = 28S coding region, p.T. = putative termination site, Z = dGdT-cluster, ⇉ = repeated regions). Region II contains all parts which are divergent in *C. th. thummi* and *C. th. piger*, including the repetitive Cla elements and the Cla-adjacent DNA. Region III represents the part that is homologous between *C. th. thummi* and *C. th. piger* and probably contains the initiation site of transcription and a major part of the ETS.

To confirm the data on the boundaries of the repetitive Cla element cluster, we repeated the sequencing work with a second, independently cloned ribosomal repeat unit of *C. th. thummi*. This confirmation seemed to be necessary, because the first investigated recombinant plasmid (pCtt 1505) was not stable during its propagation in *E. coli*. However, the sequence data obtained from the second clone (pCtt 1507) were, except for a few single base mutations, identical with the sequence data obtained from the first, unstable clone.

Comparison of the nucleotide sequence of the NTS from C. th. thummi and C. th. piger

The base sequence of the NTS-containing *Hae*III fragments of *C. th. thummi* and *C. th. piger* is shown in Figures 2 and 3. In Figure 2 the complete sequences of *C. th. piger* NTS and the homologous parts of the *C. th. thummi* NTS are put together. Some deletions/insertions are included to allow maximum homology. In Figure 3, the Cla element cluster including the Cla element adjacent DNA and the junctions with the original spacer sequences are demonstrated.

The comparison of the NTS sequences of the two subspecies allows the division of the NTS into three sections (Figure 4): the first section (region I) ranges from the 3′ end of the 28S coding region near the *Hae*III site up to the beginning of the first (left-most) repetitive Cla element, the second section (region II) represents the non-homologous parts including the repetitive Cla element cluster in *C. th. thummi*, while the third section (region III) covers the DNA sequence in which *C. th. thummi* and *C. th. piger* are homologous again and which probably includes the transcription initiation site and the major part of the ETS. This division of the NTS, shown in Figure 4, is based only on the criteria of structure and homology and does not necessarily have any functional significance.

Region I: dGdT cluster, subrepeats and putative termination site. NTS region I, probably containing the termination site of transcription, is characterized by a high degree of homology between *C. th. thummi* and *C. th. piger*. Close to the *Hae*III site there is a sequence of limited homology with the known termination sequences in *Drosophila*, yeast and *Xenopus* although it is not yet known whether this is really the termination sequence in *Chironomus*. About 500 bases downstream from this putative termination sequence there is a conspicuous dGdT cluster in the NTS of both subspecies, which might have the potential to form left-handed helical DNA (Z-DNA) (Hamada *et al.*, 1982). Downstream from

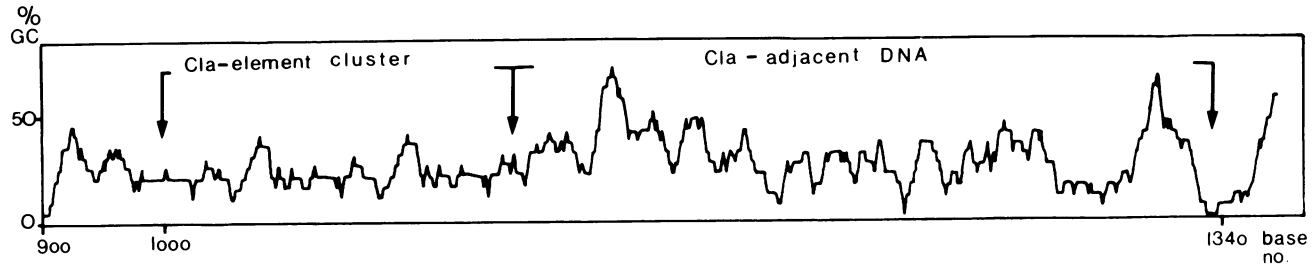

Fig. 5. Distribution of GC base pairs in region II and flanking parts of NTS of *C. th. thummi*. The GC content was calculated as the percentage of G + C pairs in a sequence of 20 bp around each nucleotide. The GC content was computer plotted against the base number as defined in Figure 2. To demonstrate the repetitive nature of the Cla element cluster, an additional Cla element was introduced into the sequence shown in Figure 3, so that 2.6 Cla elements are included in the graph. The arrows indicate the ends of the Cla element cluster, as well as the end of the Cla-adjacent DNA (approximately base 1340).

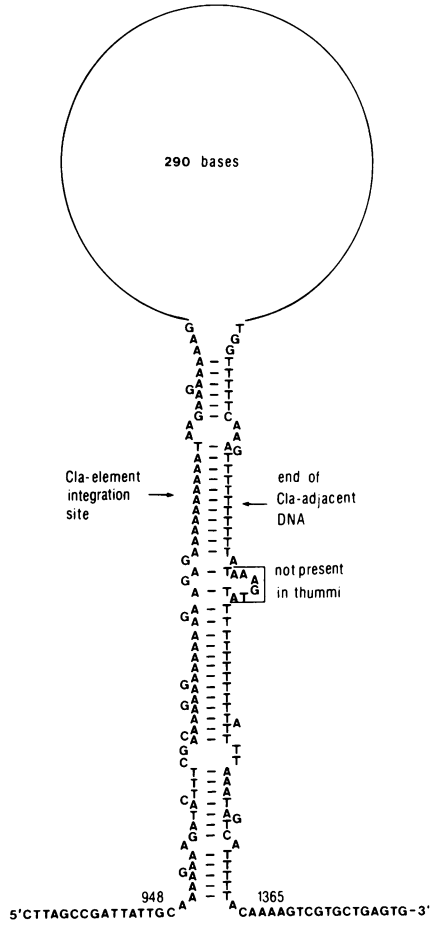

Fig. 6. Hypothetical secondary structure of the Cla element integration site. The figure shows a hypothetical intrastrand pairing of the boundary regions between region I/region II (purine tract) and region II/region III (pyrimidine tract) in the NTS of *C. th. piger*. This region is the Cla element integration site in *C. th. thummi*. The exact sites where the Cla element cluster starts and the Cla-adjacent DNA ends are indicated by arrows. The numbering corresponds to that in Figure 2.

this dGdT element, a DNA sequence which is internally repetitive starts. A piece of 65 bp (726–791) is repeated after 180 bp (906–975). The DNA between these direct repeats and upstream thereof displays some internal repetition although the possible internal subrepeats show very low homology.

Region II: highly repetitive Cla element, dA-dT clusters at the integration site. NTS region II contains the major differences of *C. th. thummi* and *C. th. piger*. The non-homologous region starts exactly with the beginning of the repetitive Cla element cluster after base 995 according to our relative numbering in Figure 2 (the numbering also includes postulated insertions/deletions in both *thummi* and *piger* NTS, so that the real number of bases deviates from that indicated by the numbering).

The first Cla element at the 5′ end of the Cla element cluster starts at base c57 relative to the *Cla*I restriction site (Figures 2, 3). This Cla element is sequenced up to the *Sau*3A1 site, which is located 33 bases downstream from the *Cla*I site, so that the *Cla*I site is included in the sequence. At the 3′ end, the Cla element cluster ends at base c9 (Figure 3) relative to the *Cla*I site, although this position is uncertain since an oligo(dT) is found at this position, which is assumed to be homologous to the Cla element despite a difference of base c5 (deletion/insertion of A).

The surprising finding that the Cla element cluster does not end with the same base as it begins with prompted us to confirm the result by sequencing this region in a second cloned rDNA of *C. th. thummi*. The nucleotide sequence of this second NTS confirmed, in principle, the result described above, although the *Cla*I site of the right-most Cla element is mutated by a single base exchange.

The repetitive Cla element cluster, however, is not the only DNA within the *C. th. thummi* NTS, which is not homologous to the *C. th. piger* NTS. Adjacent to the 3′ end of the Cla element cluster there are 528 bases which are also not present in the NTS of *C. th. piger* (Figure 3, base a1–a528). Computer hybridizations of this 'Cla-adjacent' DNA sequence to the entire *C. th. piger* NTS showed no matches above 40% homology. Thus, this sequence is not related to the original NTS as represented by *C. th. piger*. Southern hybridizations of a DNA fragment containing only this Cla-adjacent DNA-fragment with total genomic DNA of *C. th. thummi* indicate that this sequence is repetitive and widely distributed throughout the *C. th. thummi* genome (Schmidt, in preparation).

In the corresponding NTS region II of *C. th. piger* we find a 322-base DNA sequence, which has no homologous sequence in the NTS of *C. th. thummi* (Figure 2). This sequence is similar to the probably internally repetitive part of region I and might be the continuation of this region. The 322-base fragment ends at a large dT cluster, which is present also in the NTS of *C. th. thummi* (Figure 3). This dT cluster marks the beginning of region III.

In contrast to the pyrimidine tract at the junction between region II and region III, we find a 40-base purine tract at the

junction between region I and II (i.e., the 5′ end of the repetitive Cla element cluster). Both junctions are AT-rich (junction I/II = 80% AT; junction II/III = 100% AT; Figure 5). Between the junctions of region I and II and of region II and III there is a remarkably high complementarity, which might allow an intrastrand pairing exactly at the integration site of the Cla element cluster (Figure 6).

Region III: putative initiation region, ETS. Region III is defined by a high degree of homology between *C. th. thummi* and *C. th. piger*. The homology in this region (compare Figures 2 and 4) is >90%. The divergency of ~10% is mainly due to a few deletions/insertions especially in oligo(dA) clusters, which are rather frequent in this region. The transcription initiation region has not yet been mapped by S1 mapping, and a computer search for sequences homologous to the known transcription initiation region of mouse, *Drosophila* and *Xenopus* was not successful.

Discussion

The NTS of the two subspecies of *C. thummi* can be discussed as a model of evolutionary change by the transposition of a highly repetitive DNA sequence into the non-transcribed part of a gene. Because the subspecies *C. th. piger* represents the phylogenetically older one (Keyl, 1962, 1965) one can postulate that the organization of the NTS of *C. th. piger* is similar to the original NTS structure. This enables us to compare the organization of the NTS before (*piger* NTS) and after (*thummi* NTS) the transposition has occurred. The most intriguing section is of course the site where the integration of the transposed sequence took place. Furthermore, the NTS of *C. th. thummi* provides information about the ends of the repetitive Cla elements, because we can clearly determine the ends by comparing *C. th. thummi* with *C. th. piger*.

Structure of the transposed DNA sequence

The sequence determination of both ends of the repetitive Cla element cluster revealed that the left-most Cla element starts with base c57 but the right-most Cla element of the cluster stops with base c9 relative to the *Cla*I site. This result has been confirmed by sequencing a second, independently selected rDNA clone of *C. th. thummi*. Thus, the Cla element cluster consists of $n \pm 0.6$ Cla elements, and one can expect that this organization is present in all ribosomal repeat units of *C. th. thummi*.

Interestingly, the repetitive Cla element cluster does not seem to be the only DNA sequence which has been integrated into the NTS of *C. th. thummi*. The DNA sequence adjacent to the 3′ end of the repetitive Cla element cluster (extending in a 3′ direction towards the beginning of the 18S gene) seems to be part of the transposed sequence. This sequence, not internally repetitive, is not found in the corresponding region of the NTS of *C. th. piger*.

This Cla element adjacent DNA is present in many sites of the *C. th. thummi* chromosomes, probably as interspersed repetitive sequence (Schmidt, in preparation). This leads us to the conclusion, that the Cla element cluster and the Cla element adjacent DNA might have been transposed together into the ribosomal repeat units of *C. th. thummi*. Further evidence for this assumption comes from the base sequence of the integration site (see below).

Structure of the integration site

The 5′ and the 3′-flanking regions of the Cla element cluster + Cla-adjacent DNA (boundaries between region I/II and II/III) are remarkably different from the rest of the NTS sequence: at the 5′ end there is a purine tract of 23 purine bases, mainly A, in *C. th. thummi* and 40 purine bases in *C. th. piger*.

At the position where in *C. th. thummi* the Cla-adjacent DNA ends, and which is supposed to be the end of the transposed DNA, a long pyrimidine tract interrupted by some adenosyl residues is found in both subspecies. An intrastrand pairing between the poly(dA) cluster at the 5′ end of the putative transposed DNA and the poly(dT) cluster at the 3′ end of the Cla-adjacent DNA might be possible, as shown in Figure 6. The question of whether this structure has played any role during the process of transposition and integration must remain speculative, although the conspicuous base composition at these sites supports the idea that the repetitive Cla element cluster and the Cla-adjacent DNA have been transposed together.

The short terminal repeats, which result from a duplication at the integration site during the transposition (for recent review, see Shapiro and Cordell, 1982), and which is a typical feature of transposed DNAs is not present at either the ends of the repetitive Cla element cluster or at the ends of the Cla-adjacent (putative co-transposed) DNA. However, it cannot be excluded that such short terminal repeats might have existed, but were obliterated by base changes since the transposition has occurred. It is questionable whether the left end of the Cla element cluster together with the complementary right end of the Cla-adjacent DNA can be interpreted as a foldback element of the type which have been shown to exist as transposable elements in *Drosophila* (Potter *et al.*, 1980) since the length of the inverted repeat is far beyond the length reported for the foldback elements in *Drosophila*.

Another unresolved question concerns the piece of DNA of 322 bp in *C. th. piger* rDNA, which is located between the poly(dA) and the poly(dT) cluster (region II) and which is missing, or rather is replaced, by the repetitive Cla element cluster + Cla-adjacent DNA in the *C. th. thummi* rDNA. There are two possible explanations of what happened to this DNA during the transposition. The first, which we favour, assumes that this piece of DNA was deleted during the integration process of the Cla element cluster. The second possibility is that it is a transposed DNA itself, which has chosen the same habitat as the repetitive Cla element cluster + Cla-adjacent DNA, because of the pre-existing inverted repeat favouring an integration of mobile DNA sequences. The first working hypothesis is, in our view, more likely, because the repetitive Cla element cluster interrupts the poly(dA) and poly(dT) clusters (compare Figures 2 and 3) and, furthermore, these interruptions (which are identical to the site of integration) come almost at exactly the same position when the foldback structure is arranged as shown in Figure 6. In addition, region II in *C. th. piger* shows some homology to the internally repetitive portion of region I indicating that it might be a continuation of this internally repetitive region.

Origin of the repetitive Cla elements

The whole process of transposition of the highly repetitive Cla element cluster into the spacer of the rDNA of *C. th. thummi* is related to the enlargement of the genome of *C. th. thummi* relative to the genome of *C. th. piger* during its evolution (Keyl, 1965; Schmidt *et al.*, 1980; Schmidt, 1981). During the process of evolutionary DNA 'amplification', the repetitive Cla elements increased manyfold in number and oc-

cupied new positions in the chromosomes (shown by *in situ* hybridizations, Schmidt, 1981). The rRNA genes are only one example of this process, where the invasion of highly repetitive sequences into intergenic or spacer regions could be shown at the molecular level. The situation in the chironomid rRNA genes is quite similar to the situation known from the histone genes of the American Newt, *Notophthalmus viridescens*. In this animal the histone gene clusters are separated by long tracts of satellite DNA (Stephenson *et al.*, 1981; Gall *et al.*, 1981; Diaz *et al.*, 1981). This is also interpreted to be correlated to the large genome size of the newt (Stephenson *et al.*, 1981). In the case of the two chironomid subspecies this correlation is quite evident. In *C. th. piger*, the subspecies with the lower DNA content per genome, clusters of repetitive Cla elements are confined to the centromeric regions of all four chromosomes (Schmidt, 1981), while in *C. th. thummi*, the subspecies with the higher DNA content per genome, repetitive Cla elements are distributed throughout many sites in all chromosomes. Of course, we do not know whether the Cla elements found in the nucleolar DNA are of centromeric origin, but we would like to stress the high degree of homology, >90%, between the ribosomal and the consensus sequence (representing the centromeric fraction) of the *Cla*I sequence family (Schmidt *et al.*, 1982). Obviously, the integration of large amounts of this highly repetitive Cla element does not seriously interfere with the transcription of the rRNA genes, and hence the promoter function of parts of the NTS is not negatively affected. This can be postulated from the fact that most, if not all, copies of the ribosomal repeat units of *C. th. thummi* contain the repetitive Cla elements (Israelewski and Schmidt, 1982) in variable numbers, ranging from a few copies to >100 per NTS.

Although we have not yet mapped the sites of initiation and termination of polymerase I transcription, we know from R-loop mapping and Southern-hybridization using isolated 18S and 28S RNA that the beginning of the 18S gene is ~1.5 kb downstream from the repetitive Cla element cluster and ~1 kb downstream from the Cla element adjacent DNA (Schmidt *et al.*, 1982). Assuming a length of the ETS similar to *D. melanogaster* (Kohorn and Rae, 1982a) or *X. laevis* (Maden *et al.*, 1982) of 0.7–0.8 kb, then the transcription initiation site would be located 200–400 bases downstream from the end of the Cla-adjacent DNA. The site of termination would be located ~900 bases upstream of the 5′ end of the repetitive Cla element cluster.

In contrast to the organization reported for *D. melanogaster* (Kohorn and Rae, 1982b; Coen and Dover, 1982), we cannot find, by computer search, any major repetitive sequences representing a duplicated promoter region. We also do not find, by computer search, any sequences homologous to the known initiation sequences of the mouse (Bach *et al.*, 1981), *Xenopus* (Sollner-Webb and Reeder, 1979), *Drosophila* (Long *et al.*, 1981) and yeast (Klemenz and Geiduschek, 1980). Experiments to localize termination and initiation sites are in progress.

The only known consequence of the presence of the repetitive Cla elements within the NTS of *C. th. thummi* is an extraordinarily high degree of spacer length heterogeneity within the genomic ribosomal repeat units of *C. th. thummi* (Israelewski and Schmidt, 1982). This heterogeneity may be the consequence of an increased rate of unequal recombination, which is usually thought to create length heterogeneity by generating increasing numbers of repetitive elements (Smith, 1976; Dover, 1982).

Materials and methods

Plasmids

The plasmids carrying the entire ribosomal repeat units of *C. th. thummi* and *C. th. piger* were constructed as described previously (Schmidt *et al.*, 1982). For the nucleotide sequence analysis of the NTS regions we used the recombinant plasmids pCtt 1505 (*thummi* rDNA, vector pBR328), pCtt 1507 (*thummi* rDNA, vector pBR328) and pCtp 1550 (*piger* rDNA, vector pBR322).

Plasmid isolation, restriction and isolation of restriction fragments

The plasmid DNA was isolated by a rapid procedure based on the method of Birnboim and Doly (1979) with the following modifications: cells from single colonies were inoculated into 500 ml of L-broth containing 100 μg ampicillin/ml and incubated for 16 h at 37°C. Cells were pelleted and resuspended in 7.5 ml of 25% sucrose-50 mM Tris-Cl, pH 8.0. The cells were digested with lysozyme (0.4 mg/ml) at 4°C for 10 min and then lysed gently by adding 5 ml of 'Triton X-100-mix' (0.2% Triton X-100, 50 mM Tris-Cl, pH 8.0, 65 mM EDTA) and incubation at room temperature for 10 min. During this incubation, the RNA was digested with 50 μl of RNase A (10 mg/ml). After the incubation the viscous solution was centrifuged at 18 000 r.p.m. for 40 min at 0°C. The cleared supernatant was carefully removed and mixed with 0.25 volume of 5 M $NaClO_4$, extracted with Tris-saturated phenol and with chloroform-iso-amylalcohol (24:1). The pH was then adjusted to 8.5 with saturated Tris-solution and the DNA was precipitated with 2 volumes of ethanol at −20°C for at least 2 h. The DNA was collected by centrifugation, washed once with 70% ethanol, dried under vacuum and dissolved in 0.5 ml TE buffer (10 mM Tris-Cl, pH 8.0-1 mM EDTA). If the purification of the plasmid DNA from contaminating RNA fragments was required, the plasmid DNA was chromatographed on a small column (5 ml) of Bio-Gel A 15 (Bio-Rad, Munich). The plasmid DNA prepared by this method has been used for all restriction analysis, end-labelling and sequencing.

The digestion of the plasmid DNA with the different restriction endonucleases was performed as recommended by the manufacturers (*Cla*I, *Taq*I, *Eco*RI, *Hpa*II, *Bgl*II, *Xba*I – Boehringer, Mannheim; *Hind*III, *Hae*III, *Alu*I – BRL, Neu Isenburg; *Sau*3AI, *Hinf*I – New England Biolabs, Schwalbach, Ts). The restricted DNA was electrophoresed on vertical agarose slab gel using either Tris-borate (90 mM Tris, 90 mM boric acid, 1 mM EDTA) or Tris-phosphate (35 mM NaH_2PO_4, 35 mM Tris, 1 mM EDTA) as electrophoresis buffer. For preparative purposes, the ethidium bromide stained bands were cut out under u.v.-illumination and electroeluted into dialysis tubings closed by polypropylene closures (Spectrum Medical Ind., Los Angeles, CA).

End-labelling and sequencing

The DNA fragments to be sequenced were labelled either at their 3′ or 5′ end with ^{32}P using Klenow fragment of polymerase I (Boehringer, Mannheim) and [α-^{32}P]dNTP (sp. act. >3000 Ci/mmol, Amersham, Braunschweig) for 3′ end-labelling, or polynucleotide kinase (Boehringer, Mannheim or BRL, Neu Isenburg) and [γ-^{32}P]ATP (Amersham, Braunschweig) for 5′ end-labelling. DNA fragments labelled at only one end were prepared either by strand separation of the complementary strand by 'large pore' polyacrylamide electrophoresis (Szalay *et al.*, 1977), or by a second digestion with an appropriate restriction endonuclease. The single end-labelled DNA fragments were then sequenced by base-specific chemical cleavage according to the method of Maxam and Gilbert (1980). The cleavage products were separated on thin (0.5 mm) polyacrylamide gels (Sanger and Coulson, 1978), and autoradiographed using Kodak Xomat AR X-ray film with or without intensifying screens (DuPont, Cronex Hi-plus). The sequence data were computer analyzed using the programs kindly provided by R.Staden (Staden, 1980) or programs written in FORTRAN by F.J.Rolofs (Ruhr-Universitaet Bochum).

Acknowledgements

We wish to thank Professor Dr.H.-G.Keyl for his interest and his comments on the manuscript. The technical assistance of M.v.Frieling-Salewski and R.Orzechowski is greatly acknowledged. We are also grateful to H.Sommerfeld for the typing of the nucleotide sequences. This work was supported by the Deutsche Forschungsgemeinschaft.

References

Arnheim,N. and Kuehn,M. (1979) *J. Mol. Biol.*, **134**, 743-765.

Arnheim,N., Seperack,P., Banerji,J., Lang,R.B., Miesfeld,R. and Marcu, K.B. (1980) *Cell*, **22**, 179-185.

Bach,R., Allet,B. and Crippa,M. (1981) *Nucleic Acids Res.*, **9**, 5311-5330.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.
Boseley,P., Moss,T., Mächler,M., Portmann,R. and Birnstiel,M. (1979) *Cell*, **17**, 19-31.
Coen,E.S. and Dover,G.A. (1982) *Nucleic Acids Res.*, **10**, 7017-7026.
Coen,E.S., Thoday,J.M. and Dover,G. (1982) *Nature*, **295**, 564-568.
Cortadas,J. and Pavon,M.C. (1982) *EMBO J.*, **1**, 1075-1080.
Diaz,M.O., Barsacchi-Pilone,G., Mahon,K.A. and Gall,J.G. (1981) *Cell*, **24**, 649-659.
Dover,G. (1982) *Nature*, **299**, 111-117.
Gall,J.G., Stephenson,E.C., Erba,H.P., Diaz,M.O. and Barsacchi-Pilone,G. (1981) *Chromosoma*, **84**, 159-171.
Hamada,H., Petrino,M.G. and Kakunaga,T. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6465-6469.
Israelewski,N. and Schmidt,E.R. (1982) *Nucleic Acids Res.*, **10**, 7689-7700.
Keyl,H.-G. (1962) *Chromosoma*, **13**, 464-514.
Keyl,H.-G. (1965) *Chromosoma*, **17**, 139-180.
Klemenz,R. and Geiduschek,E.P. (1980) *Nucleic Acids Res.*, **8**, 2679-2689.
Kohorn,B.D. and Rae,P.M.M. (1982a) *Proc. Natl. Acad. Sci. USA*, **79**, 1501-1505.
Kohorn,B.D. and Rae,P.M.M. (1982b) *Nucleic Acids Res.*, **10**, 6879-6886.
Kunz,W., Petersen,G., Renkawitz-Pohl,R., Glätzer,K.H. and Schäfer,M. (1981) *Chromosoma*, **83**, 145-158.
Long,E.O. and Dawid,I.B. (1980) *Annu. Rev. Biochem.*, **49**, 727-764.
Long,E.O., Rebbert,M.L. and Dawid,I.B. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1513-1517.
Maden,B.E.H., Moss,M. and Salim,M. (1982) *Nucleic Acids Res.*, **10**, 2387-2398.
Maxam,A.M. and Gilbert,W. (1980) in Colowick,S.P. and Kaplan,N.O. (eds.), *Methods in Enzymology*, Vol. **65**, Academic Press, NY, pp. 499-560.
Potter,S., Truett,M., Phillips,M. and Maher,A. (1980) *Cell*, **20**, 639-647.
Rae,P.M.M., Barnett,T. and Murtif,V.L. (1981) *Chromosoma*, **82**, 637-655.
Sanger,F. and Coulson,A.R. (1978) *FEBS Lett.*, **87**, 107-110.
Schaefer,J. and Schmidt,E.R. (1981) *Chromosoma*, **84**, 61-66.
Schäfer,M., Wyman,A.R. and White,R. (1981) *J. Mol. Biol.*, **146**, 179-200.
Schmidt,E.R. (1981) *FEBS Lett.*, **129**, 21-24.
Schmidt,E.R., Vistorin,G. and Keyl,H.-G. (1980) *Chromosoma*, **76**, 35-45.
Schmidt,E.R., Godwin,E.A., Keyl,H.-G. and Israelewski,N. (1982) *Chromosoma*, **87**, 389-407.
Shapiro,J.A. and Cordell,B. (1982) *Biol. Cell.*, **43**, 31-54.
Smith,G.P. (1976) *Science (Wash.)*, **191**, 528-535.
Sollner-Webb,B. and Reeder,R.H. (1979) *Cell*, **18**, 485-499.
Staden,R. (1980) *Nucleic Acids Res.*, **8**, 3673-3694.
Stephenson,E.C., Erba,H.P. and Gall,J.G. (1981) *Cell*, **24**, 639-647.
Szalay,A.A., Grohmann,K. and Sinsheimer,R.L. (1977) *Nucleic Acids Res.*, **4**, 1569-1578.