Interspersed centromeric element with a CENP-B box-like motif in *Chironomus pallidivittatus*

Casimiro C. López and Jan-Erik Edström*

Department of Genetics, Lund University, Sölvegatan 29, S-22362 Lund, Sweden

Received May 30, 1998; Revised and Accepted July 24, 1998 DDBJ/EMBL/GenBank accession nos AF043743–AF0437746

**ABSTRACT**

Short mobile elements are present in different recom-bined forms as interspersed GC-rich islands between AT rich tandem repeats in the dipteran *Chironomus pallidivittatus*. The basic element is 80 bp long, has a pronounced invert repeat structure and contains a 17 bp segment similar to the CENP-B box in mammals. The element inserts into a specific site of the 155 bp repeat in a defined orientation surrounded by 2 bp direct repeats. The total number per genome of the main variant is <20. Elements can be present in all centromeres from *C.pallidivittatus* and the sibling species *Chironomus tentans* with pronounced differences in distribution within and between species.

**INTRODUCTION**

Higher eukaryotes have complex centromeres, often with megabase amounts of DNA (1,2), most of which is tandemly arranged and usually AT-rich (3). The role of such tandem repeats, most intensely studied in the case of the human 171 bp α satellite repeat (4), is still unclear. One particularly interesting feature of the 171 bp repeats is that subsets of them contain a 17 bp motif, the CENP-B box, binding the scleroderma antigen CENP-B (5,6). The CENP-B protein is related to transposases (7) and has been intensely studied in the case of the human 171 bp α satellite repeat (4). Isolated CENP-B repeats are GC-rich and have a highly variable intercentromeric distribution. A particularly interesting feature of this element is a sequence motif similar to the CENP-B box of mammals.

**MATERIALS AND METHODS**

Nucleotide sequence GenBank accession numbers

Cp80, AF043743; Cp136, AF043744; Cp187, AF043745; Cp223, AF043746.

Animals, DNA preparations and genomic library

Salivary glands from larvae of laboratory cultures of the two dipteran species *C.pallidivittatus* and *C.tentans* for *in situ* hybridization. Total high molecular weight genomic DNA from *C.pallidivittatus* and *C.tentans* fourth instar larvae and tissue cultured cells (20) was obtained by ultracentrifugation in a caesium chloride gradient (21) with minor modifications. DNA from individual microdissected chromosomes was isolated and treated for PCR (22). A genomic library was constructed in λZAPII (Stratagene) from *C.pallidivittatus* DNA after complete EcoRI digestion.

**PCR amplification**

PCR was in a 2400 Perkin Elmer thermocycler with buffer supplied by the manufacturer. Conditions were 8 min denaturation at 95°C, followed by 25–40 cycles, depending on template, 20 s at 94°C, 30 s at 52°C and 1 min at 72°C, with a final extension for 10 min at 72°C. The reaction was in 50 μl, with 0.2 μM of each primer, 39f and 40a (Fig. 2), 2 mM MgCl₂, 200 μM dNTPs and 1.5 U AmpliTaq Gold (Perkin Elmer).

**Probes and filter hybridizations**

PCR probes to screen the genomic library were made with primers for the insert in pCp 627 (19), which contains parts of the SINE-like element Cp1 (23) and 155 bp tandem repeats. Probes labelled by direct incorporation with PCR were made from both components. Two micro litres of each amplification reaction (~10 ng) was cycled eight times in the presence of 30 μCi [α-32P]dCTP, 0.2 μM each primer, 4.5 μM dATP, dGTP and dTTP and 0.5 U AmpliTaq (Perkin Elmer) in 50 μl. Before hybridization, labelled probes were gel filtered and incorporation measured. For genomic analysis probes representing the interspersed element Cp80 (see...
Results were labelled similarly. Hybridization conditions were 6x SSC, 0.5% SDS, 5x Denhardt’s solution and 100 μg/ml salmon sperm DNA at 65°C. Final washes were in 0.5x SSC, 0.3% SDS at 65°C.

Cloning and DNA sequencing

PCR products were blunt-end cloned at the dephosphorylated EcoRI site. Before ligation, PCR fragments were treated with T4 polymerase (Boehringer Mannheim), followed by T4 kinase (Promega). Screening of recombinants was with plasmid or pBluescript SK(–) phagemid clones. Sequencing was with T4 polymerase (Boehringer Mannheim), followed by T4 kinase (Promega). Screening of recombinants was with plasmid or pBluescript SK(–) phagemid clones. Sequencing was performed at a vector:linker molar concentration ratio of 1:5. Screening of the Cp80–MIMIC construction was by colony PCR using standard protocols. Genomic DNA and the cloned MIMIC were determined spectrophotometrically and controlled by photometry of an EtBr-stained agarose gel. In a preliminary titration a constant amount of genomic DNA (20 ng), corresponding to 100 000 genome equivalents, was added to a PCR reaction containing 10-fold serial dilutions of Cp80–MIMIC. To quantify PCR products [α-32P]dCTP was included in the reaction. After 33 and 39 cycles, 20% of the reaction was run in a 3.7% EtBr–agarose gel (Nu Sieve GTG; FMC Bioproducts). Following electrophoresis, bands corresponding to the target and MIMIC were excised from the gel and radioactivity determined by scintillation counting. (The radioactivity in the two bands was compared taking into account the cytosine content of the sequences.) For more accurate quantification, a second experiment was performed with 3-fold dilutions up to 33 genome equivalents of the MIMIC sequence.

Figure 1. Schematic representation of genomic clones obtained by differential screening with the 155 bp repeat and Cp1 sequence. The arrays of 155 bp repeat are indicated with one arrow for each repeat. Adjoining boxes terminating with an EcoRI site show the 36 and 44 bp sequences that compose the Cp80 repeat. The size of DNA fragments which are artefactually ligated to the EcoRI sites during cloning is indicated. E, EcoRI; S, SacII.

Cloning and DNA sequencing

PCR products were blunt-end cloned at the dephosphorylated pUC18 Smal site. Before ligation, PCR fragments were treated with T4 polymerase (Boehringer Mannheim), followed by T4 kinase (Promega). Screening of recombinants was with plasmid primers and by colony PCR. Dideoxy sequencing was on pUC18 plasmid or pBluescript SK(–) phagemid clones. Sequencing was manual with an AmpliCycle Sequencing Kit (Perkin Elmer) or with an automatic sequenator (310 DNA Sequencing System; Applied Biosystems). Results were analysed with the University of Wisconsin GCG Software Package. DNA sequences were compared with databases using BLAST (24).

Fluorescence in situ hybridization

Probes were biotin-labelled by PCR with a nucleotide mixture containing 4 μM biotin-16-dUTP, 4.5 μM dATP, dCTP and dGTP and 0.45 μM dTTP. Squashes were from salivary glands and hybridization was overnight at 58°C. For detection a three-step method was used with avidin–fluorescein isothiocyanate conjugate (Sigma) applied before and after biotinylated anti-avidin D (Vector Laboratories).

Quantification of sequences by competitive PCR

Genomic amounts of Cp80 sequence were estimated by competitive PCR. The MIMIC target was generated by insertion of a 23 bp fragment at the SacII site. Target site duplications are boxed and in bold. The putative CENP-B box is underlined. Palindromes, pd1 and pd2, are indicated by arrow pairs.

RESULTS

Isolation of centromere-associated sequences

The λZAPII genomic library from C.pallidivittatus, screened with the 155 bp repeat, gave 30 positive clones. In a second screen seven clones hybridizing with Cp1 components were eliminated. Five of the 23 remaining clones were selected for further analysis. All inserts had a similar organization (Fig. 1). A variable number of 155 bp tandem repeats were flanked by the same 36 bp sequence in the 5'-direction, ending with the EcoRI site, and all of them had the same 44 bp in the 3'-direction of the 155 bp repeats, up to the terminating EcoRI site. Some of the clones contained additional DNA outside this site, likely arising from ligation of unrelated restriction fragments during library construction. Only the size of these segments is therefore indicated in Figure 1. The 36 and 44 bp fragments of all clones show only one base difference in comparisons between corresponding sequences (C→T at position 101, Fig. 2). The two kinds of fragments join 155 bp repeats from opposite ends so that an integral number of 155 bp repeats followed by 2 bp (CA), identical to the start of another 155 bp repeat, become interposed between the 36 and 44 bp fragments.

Sequence of the basic 80 bp unit

PCR was performed on genomic DNA with primers hybridizing close to the junctions of the fragments with the 155 bp repeats and a 72 bp band was cloned and sequenced (Fig. 2). The 72 bp segment contains most of the 36 and 44 bp fragments without intervening DNA and the expected EcoRI site is present. The
derived unit, termed Cp80 (the primers exclude 8 bp at the ends), is surrounded by 2 bp direct repeats (CA) from the 155 bp repeat. It contains two 14 bp overlapping palindromes, pd1 and pd2. The central part of pd1 is a PstI restriction site, whereas pd2 contains a SacII site. There is also a 9 bp direct interspersed repeat (positions 38–46 and 71–79), which becomes 19 bp long (A TTCc/aGCG-GAA TTAa/tTTTC) if two deviations are allowed (Fig. 2). Close to its left end Cp80 contains a 17 bp sequence (TTTCGGAAA TT-CAGCGGAA) similar to the CENP-B box consensus (25). Insertion of Cp80 is in the AT-rich more conserved part of the 155 bp repeat (26) and occurs 25 bp upstream of the HindIII site characterizing this repeat. With 10-fold increased genomic template a 128 bp PCR band appeared, Cp136, a complex unit with irregular triplication of pd2 and duplication of most of pd1, also containing two short sequences of unknown origin.

**In situ localization of Cp80 sequences**

Cp80 hybridized in situ exclusively to centromere regions of *C.pallidivittatus* (Fig. 3A and B) and the sibling species *C.tentans* (Fig. 3C) with distinct differences in intercentromeric distribution. In a given animal not all centromeres were as a rule visibly labelled. Figure 3A shows an animal where hybridization is restricted to centromeres 2 and 4 and Figure 3B another one where it is seen only over centromere 2. In Figure 3C centromeres 1–3 show weak but clear signals.

**PCR amplification of the Cp80 sequence from individual chromosomes**

Since Cp80 represents a short target, only clustered elements are likely to be detected in situ. For a more sensitive assay we pooled chromosomes from fixed salivary glands of *C.pallidivittatus*, 10 of each kind, for PCR, after which the original amplification products were re-amplified. Intense spots of Cp80-hybridizing material were obtained from chromosomes 2 and 4 and weaker bands from chromosomes 1 and 3. In addition to the already cloned 80 and 136 bp units, two new products were seen at 179 and 215 bp, the largest of which was restricted to chromosomes 1 and 3 which did not contain any other products (Fig. 4A). The bands were re-amplified and sequenced and the products designated Cp187 and Cp223. In Cp187 a piece of 155 bp repeat is inserted between one complete and another incomplete Cp80 element. Cp223 contains three central Cp80 segments in tandem preceded and followed by Cp80 units truncated at their 3′- and 5′- ends by 24 and 37 bp, respectively (Fig. 4B). The three central Cp80 components each contain the whole of pd2 and the preceding 5 bp and they are all surrounded by an unrelated 14 bp sequence that in one case is extended by another 11 bp of unknown origin. All in all, there are five pd2 palindromes in Cp223, one of which has a single base mutation. Cp223 thus resembles Cp136 in having additional pd2-containing segments, although here surrounded by unrelated DNA. Cp187 may have been formed from two originally complete Cp80 units separated by several 155 bp
Figure 5. Genomic organization of the Cp80 sequence. (A) Southern blot of genomic DNA hybridized with labelled Cp80 sequence after separation in a 0.7% agarose gel. Digestion with HindIII (h) shows low molecular weight signals and a 4 kb band (each indicated by a dot). EcoRI (e) distributes material with reduced hybridization (the enzyme cutting in the central part of Cp80) over a broad range, whereas SacI (s1), BamHI (b) and AccI (a) give high molecular weight hybridizing DNA. (B) Digestion with PstI (p) and SacII (s2) and separation in a 0.4% agarose gel followed by blotting and hybridization with labelled 155 bp repeat. m, 5 kb size ladder.

repeats followed by deletion of all except part of a 155 bp repeat and part of a Cp80 element. Cp80, Cp136 and Cp223 each contain one CENP-B box-like motif, whereas there are two such motifs in Cp187.

Genomic organization of the Cp80 sequence

Digestion of genomic DNA with HindIII, cutting once per 155 bp repeat, produced three bands hybridizing with Cp80 (Fig. 5A), likely to be a result of 80, 136 and 187 bp units added to the 155 bp sequence (expected sizes, including the CA direct repeat, 237, 293 and 344 bp). In addition, a band of ~4 kb of unknown origin was found, possibly representing a Cp80-containing fragment including DNA flanking a repeat array. Digestion with EcoRI, which cuts Cp80 in the central part, decreased its hybridization and produced a broad size range of weakly hybridizing fragments. Most enzymes with six base recognition specificities (three of which are shown in Fig. 5A) gave signals at the mobility limit of the gel, confirming that Cp80 and derivatives are associated with repetitive DNA.

Genomic DNA was digested with either SacII or PstI, with sites in Cp80, separated in a 0.4% agarose gel, blotted and hybridized with the 155 bp repeat (Fig. 5B). Most DNA remained of large size after digestion, but many bands of smaller size were created, the majority of which had the same size after digestion with both enzymes, which shows that the majority of Cp80 elements contain both restriction sites. Most fragments are likely not to contain complex DNA flanking the arrays of 155 bp repeats, since this would probably have given different sizes of fragments after cutting with the two enzymes. Therefore, the Cp80 family of elements are within long tandem blocks of 155 bp repeats, although some element(s) might be close to the end of such arrays.

Genomic abundance of Cp80-derived sequences

The number of Cp80 copies per genome determined by competitive PCR was in the range 14–18 (Fig. 6). There was no significant difference between their abundance in DNA from larval, predominantly polytene tissue in the two sibling species C.pallidivittatus and C.tentans (Fig. 6A and B), with similar values also for DNA from actively dividing diploid cells in C.tentans (Fig. 6C).

DISCUSSION

The 155 bp centromeric repeat arrays in C.pallidivittatus contain members of an interspersed family of elements, i.e. a predominant 80 bp unit (Cp80) and three recombined versions, 136, 187 and 223 bp long. Cp80 is strongly palindromic and contains a sequence motif similar to the human CENP-B box. Analyses of cloned units and blots of genomic DNA suggested that most elements are interspersed as relatively GC-rich islands between the AT-rich 155 bp repeats. There are somewhat less than 20 copies of the predominant form in DNA from diploid cells (tissue culture) and from larval, largely polytene tissue.

Cp80 is likely to be mobile, being surrounded by short sequence duplications. Furthermore, its distribution is highly
variable within and between species. Since it cannot code for protein, transposition would require extraneous enzymes. In spite of its palindromic structure, it does not resemble any known DNA transposon-like elements which can, however, be as small as Cp80 (7). Cp80 also differs from SINEs, among other things in not having internal polymerase III internal control elements (27,28). It is also unlikely to be a pseudogene, representing an mRNA 3'-end. Another mobile element of small size, designated Cpi, is present in Chironomus centromeres (29), but also in the extracentromeric genome (23). As for Cp80, a molecular basis for mobility is unknown.

Even if the function of Cp80 has not been elucidated, its structure and distribution may give useful hints about a possible cellular role. Of particular interest is a 17 bp motif in the left half of the CENP-B box consensus (YTTCGT-TGGAARCGGGA) (25). Only 2 of the 9 bp essential for binding to the CENP-B protein (bold) disagree in the anti-CENP-B protein (bold) and the CENP-B box consensus (lower case). Attempts to bind human CENP-B protein from a HeLa nuclear extract (as assayed with the aid of anti-CENP-B protein) to the CENP-B box consensus (underlined) to the CENP-B box consensus (Y

REFERENCES


ACKNOWLEDGEMENTS

The present investigation was supported by grants from the Swedish Cancer Society, the Nilsson-Ehle Foundation, the Philip-Sörensén Foundation and the Crafoord Foundation. We are also indebted to Dr W. C. Earnshaw (Edinburgh) for the gift of CENP-B antibody and to Dr Irene Kannert for revising the manuscript.