3728–3736 Nucleic Acids Research, 1994, Vol. 22, No. 18

Cloning and molecular characterization of a novel chromosome specific centromere sequence of Chinese hamster

Károly Fátyol*, Imre Cserpán, Tünde Praznovszky, Judit Kereső and Gyula Hadlaczky
Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, H-6701 Szeged, PO Box 521, Hungary

Received May 31, 1994; Revised and Accepted July 25, 1994 EMBL accession nos X79295–X79299 (incl.)

ABSTRACT

We isolated and characterized the first chromosome-specific satellite DNA (HC2sat) of Chinese hamster. This novel satellite was localized to the pericentric region of hamster chromosome 2. The 2.8 kb long repeat unit of HC2sat was identified and two such units were sequenced. Extended short range periodicity could not be revealed in repeat units. These elements are amongst the largest satellite repeat units reported from mammals to date. HC2sat is a major constituent of the pericentric region of CHO chromosome 2 representing a 7–14 Mb long DNA segment. Studies of long range organization of HC2sat indicated that the formation of the satellite array might occur in different phases and involved different amplification mechanisms.

INTRODUCTION

The genome of higher eukaryotes contains large amount of seemingly useless interspersed and tandemly repeated sequence elements. Although our knowledge regarding the function of repetitive sequences is limited, there are indications for their involvement in organization and function of chromatin (1). Interspersed repeats are distributed more or less evenly in the genome, however, recent studies have revealed the concentration of different types of repeats in distinct chromosomal bands (2). The origin of interspersed repeats now has become clear, however, their role in the present-day eukaryotic genome is enigmatic. Interspersed repeats might contribute to some general aspects of eukaryotic chromatin: perhaps they act as replication origins and may play a role in the fluidity of the genome (3, 4, 5).

Satellite sequences are characterized by DNA segments (repeat units) repeated up to several thousand times in a tandem manner. They usually reside on the telomeric and pericentric regions of chromosomes. The characteristic association of satellites with these heterochromatic, structurally resistant chromosome regions indicates their possible role in the organization and function of telomeres and centromeres.

DNA elements necessary for protection of chromosome ends have been identified. A terminally located DNA fragment, containing an ~0.5 kb $(TTAGGG)_n$ sequence is necessary and sufficient for *de novo* telomere formation in mammalian cells (6). Subtelomeric repeats are thought to have only minor effects on telomere structure (7).

DNA sequences capable of forming a centromere have not yet been identified in mammalian cells, although, there are observations that tandemly organized satellite sequences located near centromeres affect chromosome segregation (9, 10).

In human chromosomes a number of satellite DNA families have been identified. Their organization and possible function have been studied in detail. Two of them, the alpha satellite and perhaps pentameric satellite are present at all human centromeres (7). It has been shown that repeated units of alpha satellites, containing a 17 bp long sequence motif (CENP-B box) can interact with an evolutionarily conserved centromere protein (CENP-B) (11, 12). Although CENP-B is one of the major protein components of most human centromeres it could not be detected on the Y chromosome. This fact, and the stability of human chromosomes in hamster cells from which the CENP-B box is absent (19), raise questions about the role of the alpha satellite in centromere function. Recent descriptions of a natural chromosome 14 variant and a marker chromosome 10, containing no detectable amounts of either pentameric or alpha satellites (13, 14), also make the role of these sequence elements in proper chromosome segregation uncertain.

A number of different chromosome specific repetitive sequences have been described. In human, alpha satellite subfamilies specific to several chromosomes have been identified. Chromosome specificity of alpha satellites is mainly based on sequence divergence of repeat units but specific higher order repeat structures can also contribute to the characteristic organization of a repeat array (15).

Repetitive sequences specific to mouse chromosome X and 8 have also been isolated. Both repetitive elements are present in relatively low copy number on a given chromosome and are parts of long complex repeat elements (16, 17).

*To whom correspondence should be addressed

Chromosome specific repetitive sequences have not been known in Chinese hamster. Centromeric regions of most hamster chromosomes comprise TTAGGG telomeric hexanucleotides. The pericentric regions of large hamster chromosomes (#1 and #2) are free from large blocks of telomeric repeats (18). Aside from a recently described 6 kb long single copy DNA fragment specific to chromosome 1 (19), centromeric DNA has not yet been isolated from these chromosomes.

In this paper we report the isolation, cloning and molecular characterization of the first chromosome specific centromeric satellite sequence of Chinese hamster. To isolate centromeric DNA fragments from Chinese hamster chromosomes we applied the *in vivo* chromosome fragmentation technique described recently (19), combined with immunoprecipitation using anti-centromere serum. This novel satellite is located exclusively on chromosome 2. The identified 2.8 kb long repeat unit is one of the longest homogeneous repetitive elements reported from mammals to date. An estimated 2,500–5,000 copies of repeat units occupy a 7–14 Mbp long chromosome segment making the novel satellite the major component of the pericentric region of chromosome 2.

MATERIALS AND METHODS

Cells and cell lines

HeLa, LMTK$^-$ and CHO-K20 cells were grown in F12 medium supplemented with 10% FCS. Concanavalin A (5–10 mg/ml)-induced Chinese hamster spleen cells were grown in RPMI medium for 72 h.

In vivo chromosome fragmentation, immunoprecipitation of chromosome fragments and plasmid sublibrary construction

4×10^8 CHO K-20 cells were cultured for 20 h in F12 medium in the presence of 2 mM hydroxyurea. The medium was supplemented with 5 mM caffeine and 0.5 mg/ml colchicine and cells were incubated for additional an 5.5 h. Mitotic cells were harvested by selective detachment and collected by centrifugation (200 *g* for 10 min). Pelleted cells were resuspended in glycine–hexyleneglycol–Triton X-100 (GHT) buffer (21), supplemented with 1 mM PMSF and burst by passing through a 23 gauge needle several times. Nuclei, cell debris, intact chromosomes and large chromosome fragments were removed by differential centrifugation (1500 *g* for 20 min) and supernatant was collected.

400 mg Protein-A–Sepharose CL4B (Pharmacia) was swollen in sterile PBS containing 0.05% Tween-20 (Tween–PBS) and incubated for 2 h at room temperature with 100 ml of human anti-centromere serum LU 851 (21). Sepharose was washed with Tween–PBS, mixed with supernatant and incubated for 3 h at room temperature with gentle agitation. Sepharose was pelleted by low speed centrifugation and washed three times (5 min each) with Tween–PBS. Pellet was resuspended in 3 ml of TE buffer containing 1% SDS, 100 mg/ml Proteinase-K and incubated for 8 h at 50°C. Immunoprecipitated DNA was phenol extracted, precipitated with isopropanol in the presence of 0.2 M ammonium acetate and washed with ethanol. Purified DNA was digested with *Nae*I restriction endonuclease, end-repaired with Klenow polymerase and cloned into pUC 19 vector.

DNA techniques, sequencing

All DNA manipulations were performed using conventional techniques (31). DNA probes used for Southern hybridization were labeled using Megaprime kit (Amersham). The CHO-K1 genomic lambda phage library (Stratagene) was screened as described (31). Di-deoxy sequencing reactions were performed on single stranded M13mp18 or M13mp19 phage clones. Sequence analyses were carried out manually using Sequenase (USB) or with an automatic sequencer (373 DNA Sequencing System, Applied Biosystems).

Sequence analyses, homology comparisons

Nucleotide sequence analyses were carried out using MicroGenie (Beckman) and PC/Gene programs. DNA sequences were compared with the GenBank (Release 82.0, April 9, 1994) and EMBL (Release 38.0, April, 1994) databases using BLAST (32).

Pulsed field gel electrophoresis

For preparation of high molecular weight DNA 2×10^8 CHO-K20 cells were harvested, washed with PBS and resuspended in 2.5 ml PBS. An equal volume of 1% LMP agarose in PBS was added and the cell suspension was poured into block moulds. After chilling, the blocks were incubated in ESP buffer (0.5 M EDTA, pH 8, 1% Na-lauroyl-sarcosine; 100 mg/ml Proteinase-K) at 50°C overnight. Blocks were stored in 50 mM EDTA (pH 8) at 4°C.

Prior to restriction enzyme digestions blocks were incubated in TE buffer containing 1 mM PMSF for 2 h at 37°C, washed with TE buffer and equilibrated with recommended restriction buffers. Restriction enzyme digestions were carried out according to the manufacturer's instructions. PFGE was performed on a home-made 'waltzer' (rotating gel system) apparatus. Restriction enzyme digested high molecular weight DNA was separated in 1% agarose gel in 0.25×TBE buffer at 180 V. The temperature was maintained at 15°C for 30 h. The pulse time was 20 s and the angle of rotation was set at 110°.

In situ hybridization

DNA probes were labelled with biotin 11–dUTP (Sigma) by nick translation using the Amersham nick translation kit. *In situ* hybridizations were carried out as described (8).

Microscopy

Observations and microphotography were made by using an Olympus AHBS Vanox microscope. Fujicolor 400 Super HG color films were used for photographs.

RESULTS

Isolation of a chromosome specific centromere sequence from CHO cells

Hydroxyurea treatment of CHO cells inhibits DNA synthesis and arrests cells at the G1/S boundary. Addition of caffeine to hydroxyurea-treated cells (HUCAFF treatment) induces mitosis coupled with premature condensation of unreplicated chromatin. It has been shown that centromeric chromosome fragments attach to microtubules of the mitotic spindle, align on the equator of the dividing cell and show apparently normal anaphase movements. These observations indicate that hydroxyurea–caffeine induced *in vivo* chromosome fragmentation has no considerable damaging effect on centromere function (20).

We induced chromosome fragmentation in CHO cells by HUCAFF–colchicine treatment. This procedure generated chromosome fragments free of mitotic microtubules. Chromosome fragments were liberated from cells by detergent

lysis and small chromatin pieces were separated from nuclei and intact chromosomes by differential centrifugation. Chromosome fragments carrying centromeric proteins were immunoprecipitated from the supernatant by using a human anti-centromere serum (21). DNA molecules were purified from immunoprecipitated material, and used to produce a plasmid sublibrary.

Hamster DNA fragments were excised from 8 randomly selected plasmid clones and used as probes for *in situ* hybridization to metaphase CHO chromosomes. One clone (HUCAFF170) showed strong hybridization to the centromeric regions of two hamster chromosomes (data not shown). To identify the hybridizing pair of chromosomes we performed *in situ* hybridization on Chinese hamster spleen cells. In this way, we were able to determine that the cloned DNA fragment originated from the pericentric region of hamster chromosome 2 (Fig. 1).

The isolated centromeric DNA is part of repeat units of a novel chromosome specific satellite

The sequence of the 536 bp HUCAFF170 insert showed no extended homology to known DNA sequences. The species specificity of the cloned DNA fragment was examined by *in situ* and Southern hybridization experiments. The fragment did not hybridize to metaphase human and mouse chromosomes. In Southern hybridization HUCAFF170 clone found no homologous sequences in human, mouse or Syrian hamster genomic DNA under stringent conditions. These results indicated that the novel centromeric sequence was evolutionarily not conserved, and was specific to the Chinese hamster genome.

The intensity of the *in situ* hybridization signal obtained with the HUCAFF170 clone indicated that the DNA segments homologous to the 536 bp DNA fragment were present in high copy number in the centromeric region of chromosome 2. To characterize the organization of this genomic region, we isolated a lambda phage clone (HC90) from a Chinese hamster genomic library by using the 536 bp fragment as a probe. The 13.3 kb insert of HC90 phage hybridized to the centromeric region of chromosome 2 (data not shown).

Restriction mapping revealed that the 13.3 kb HC90 insert contained three tandemly arranged 2.8 kb *Sac*I fragments hybridizing to HUCAFF170 (Fig. 2a). Two of them showed identical restriction patterns. Both units contained a 1.8 kb *Bgl*II fragment at the same position. We subcloned the *Bgl*II fragment in a plasmid vector (pHC450, Fig. 2a) and used it as a hybridization probe to metaphase hamster chromosomes. The 1.8 kb *Bgl*II fragment hybridized to chromosome 2 similarly to HUCAFF170 and HC90 (data not shown). The third 2.8 kb *Sac*I fragment separating two identical *Sac*I units showed a different restriction pattern. It contained a 2.4 kb *Bgl*II−*Sac*I subfragment cross-hybridizing to pHC450 and HUCAFF170 clones.

From these results we concluded that the tandemly arranged 2.8 kb related *Sac*I fragments were repeat units of a novel satellite. The originally isolated 536 bp HUCAFF170 fragment was homologous to the 3′ ends of these units. The two different *Sac*I fragments of HC90 phage might represent two repeat unit variants (type **'a'** and type **'b'**, Fig. 2b,c) of the satellite.

HC90 clone contained a 1.2 kb *Bam*HI−*Sac*I fragment adjacent to the type **'a'** unit. Since this fragment hybridized to HUCAFF170, it could be regarded as a mutilated repeat unit (Fig. 2b, see later). The leftmost 3.2 kb *Sal*I−*Bam*HI fragment had no homology to *Sac*I repeat units. We subcloned the extreme left part of this region in a plasmid vector (pHC312, Fig. 2a) and used it for *in situ* hybridization to metaphase hamster chromosomes. The fragment hybridized evenly to arms of all hamster chromosomes, indicating the presence of an interspersed repeat (data not shown). Sequence analyses confirmed that this fragment indeed contained a LINE element (unpublished result).

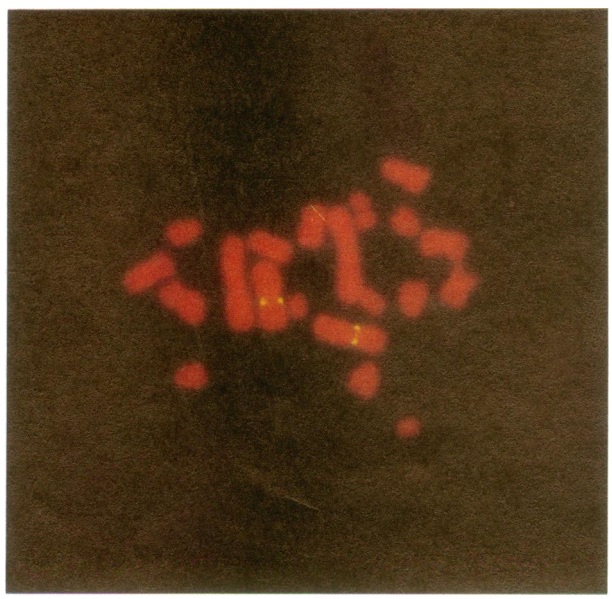

Figure 1. *In situ* hybridization of biotin-labeled HUCAFF170 probe to Chinese hamster chromosomes. Metaphase spreads were prepared from spleen cells. Chromosomes were counterstained with propidium iodide.

Nucleotide sequence of type 'a' and type 'b' repeat units

The 2.8 kb *Sac*I fragments corresponding to type **'a'** and type **'b'** repeat units were subcloned in M13 vectors and nucleotide sequences were determined. The 2,847 bp type **'a'** unit showed 93.6% homology to the 2,843 bp type **'b'** unit. In a computer search, neither of the repeat units found extensive homology in databases. Internal short range periodicity spread over the whole *Sac*I units could not be found. The overall base composition of repeat units were symmetrical (G−C: 52.1%). The central regions were A−T rich, while 5′ and 3′ ends showed higher G−C concentrations (data not shown).

In central parts of the repeat units we found some sequence peculiarities that are generally thought to be associated with genomic instabilities (26). These regions contained different simple repeat tracts and subregions with high A−T content (Figure 2c). The type **'a'** unit contained an $(ATTT)_9$ tract. In one out of nine tetranucleotide repeats there was a T→C transition. In the corresponding region of the type **'b'** unit the tandemly reiterated ATTT tetranucleotide repeats were replaced by ATTG repeats. Two out of six repeats contained T→A transversions. The simple repeat tracts are flanked by pyrimidine rich degenerated direct repeats in both units.

Close to the A−T rich simple repeats there were long $(CA)_n$ tracts. The ends of the $(CA)_n$ tracts showed sequence variations, which was consistent with Smith's predictions made for tandemly repeated arrays generated by unequal exchanges (22).

The 400 bp long fragments (1,100−1,500) of repeat units connecting two simple repeat tracts contained A−T rich regions, $d(A-T)_{n>3}$ runs and short sequence motifs repeated two or three times in a tandem manner.

The sequence differences between '**a**' and '**b**' units showed non-random distribution (Fig. 2d). The 5′ and 3′ ends of the repeat units had more than 99% homology. On the other hand the homology between the central regions was only 88.5%. The mutations were mainly transitions and transversions but deletions and insertions also occurred close to the simple repeat tracts. A possible explanation for this significant divergence is that the central part of an '**a**' unit had been replaced with sequences of a diverged repeat variant by unequal double crossing over, which resulted in the formation of the unit '**b**'.

The 369 bp 5′ segment of the HUCAFF170 insert had 90.6% homology to the corresponding regions of the *Sac*I units. The 3′ end was unrelated to these repeat units and did not show homology to known DNA sequences. Whether this junction was present in the genome or it was generated during cloning remains to be elucidated.

The 1.2 kb long *Bam*HI−*Sac*I fragment is a mutilated repeat unit

The 1.2 kb long *Bam*HI−*Sac*I subfragment of HC90 phage hybridizing the to HUCAFF170 insert was subcloned (1.2BS, Fig. 2a) and its nucleotide sequence was determined. The 1,164 bp fragment consisted of two unrelated regions. The 617 bp segment adjacent to the type '**a**' repeat unit showed 99.2% and 98.2% homology to 3′ ends of type '**a**' and '**b**' units, respectively. The A−T rich 5′ end of the fragment had no homology to the repeat units. These two parts were separated by a 66 bp segment, which showed 86.4% homology to the 130−195 regions of the *Sac*I units. These observations indicated that the 3′ end of the fragment corresponded to a mutilated repeat unit from which a 2.1 kb long segment had been deleted (Fig. 2b). The possibility that this junction was a cloning artifact generated during ligation

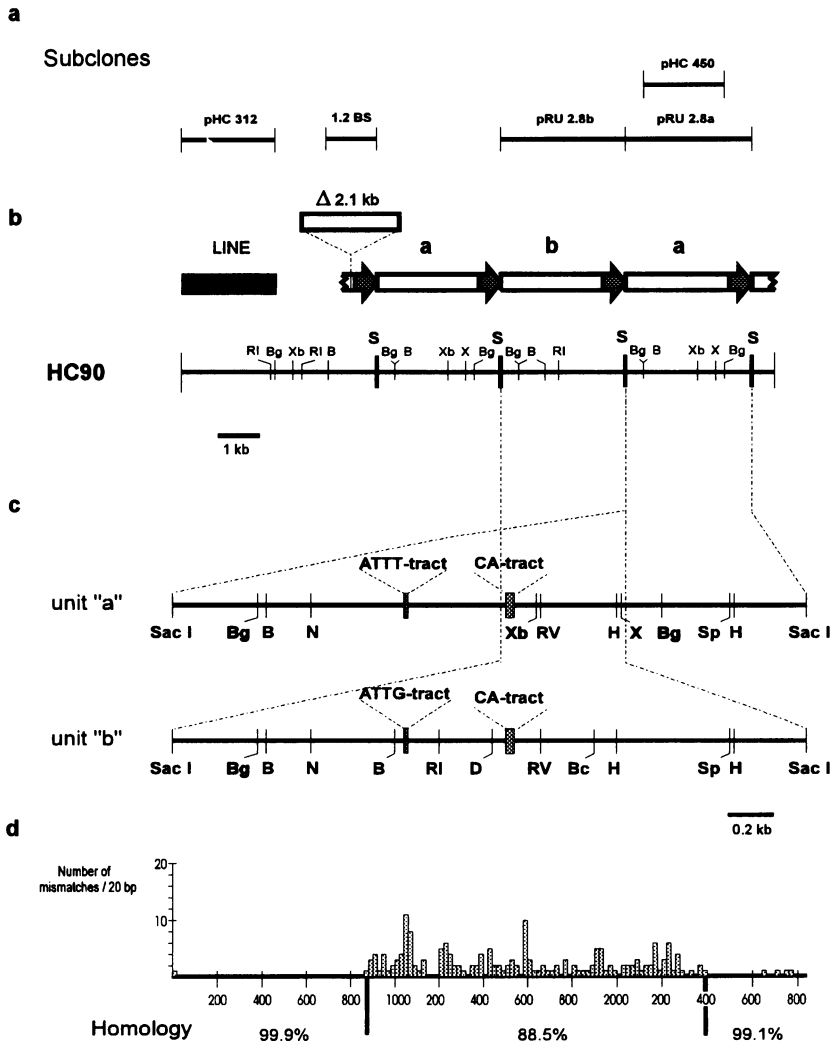

Figure 2. (**a**) Subclones of HC90 phage used in these studies. Indicated fragments were subcloned into pUC18 or pNEB193 plasmid vectors. (**b**) Restriction map of HC90 phage. Arrows indicate tandemly organized 2.8 kb repeat units. Shaded regions of repeat units indicate segments homologous to HUCAFF170. Black box corresponds to a LINE element. (**c**) Restriction maps of type '**a**' and '**b**' repeat units. (**d**) Distribution of sequence differences between '**a**' and '**b**' repeat units. Abbreviations of restriction endonucleases: B: *Bam*HI, Bc: *Bcl*I, Bg: *Bgl*II, D: *Dra*I, H: *Hin*dIII, N: *Nco*I, RI: *Eco*RI, RV: *Eco*RV, S: *Sac*I, Sp: *Sph*I, X: *Xho*I, Xb: *Xba*I.

of *Sau*3A fragments into the phage vector can be ruled out by the absence of *Sau*3A sites from the region connecting the two unrelated sequences.

Copy number of the novel repetitive sequence

Known amounts of subcloned 2.8 kb *Sac*I unit and genomic hamster DNA were blotted onto nitrocellulose filters and hybridized with radioactively labeled fragments of the repeat unit. Hybridization intensity of genomic DNA was compared to hybridization signals obtained on different amounts of cloned *Sac*I fragment. By this method the copy number of the novel satellite was estimated to be 2,500−5,000 per haploid genome. Consequently, the hamster chromosome 2 satellite (HC2sat) sequences occupy a 7−14 Mb region of chromosome 2.

Genomic organization of HC2sat sequence

To investigate the genomic organization of the novel repetitive sequence, restriction enzyme digested Chinese hamster DNA was separated by conventional or pulsed field gel electrophoresis and hybridized to radioactively labeled DNA fragments. We used different probes in these experiments: subcloned 2.8 kb *Sac*I fragments (pRU2.8a, pRU2.8b), HUCAFF170 and pHC450 inserts (Fig. 2a).

Figure 3 shows results of Southern hybridization experiments. The hybridization patterns obtained with the HUCAFF170 (Fig. 3a) and pHC450 (Fig. 3b) probes were very similar, which indicated that DNA segments homologous to the probes were linked in the genome. The *Sac*I, *Bgl*II and *Hin*dIII sites were located at an apparently identical position in the majority of genomic repeats, like in types '**a**' and '**b**' units as confirmed by double digests (Fig. 3c).

In the majority of repeat units *Sac*I, *Nco*I, *Sph*I, *Stu*I and *Nde*I cleaved only once. The relative positions of cleavage sites were similar to that observed in types '**a**' and '**b**' repeat units (Fig. 3d).

An intense 2.8 kb band appeared in all restriction digests and the ladder-like hybridization patterns of *Nde*I and *Eco*RI digested DNA confirmed the tandem organization of the 2.8 kb repeat units (Fig. 3e). The weak bands observed were thought to represent less abundant repeat unit variants and/or junction regions.

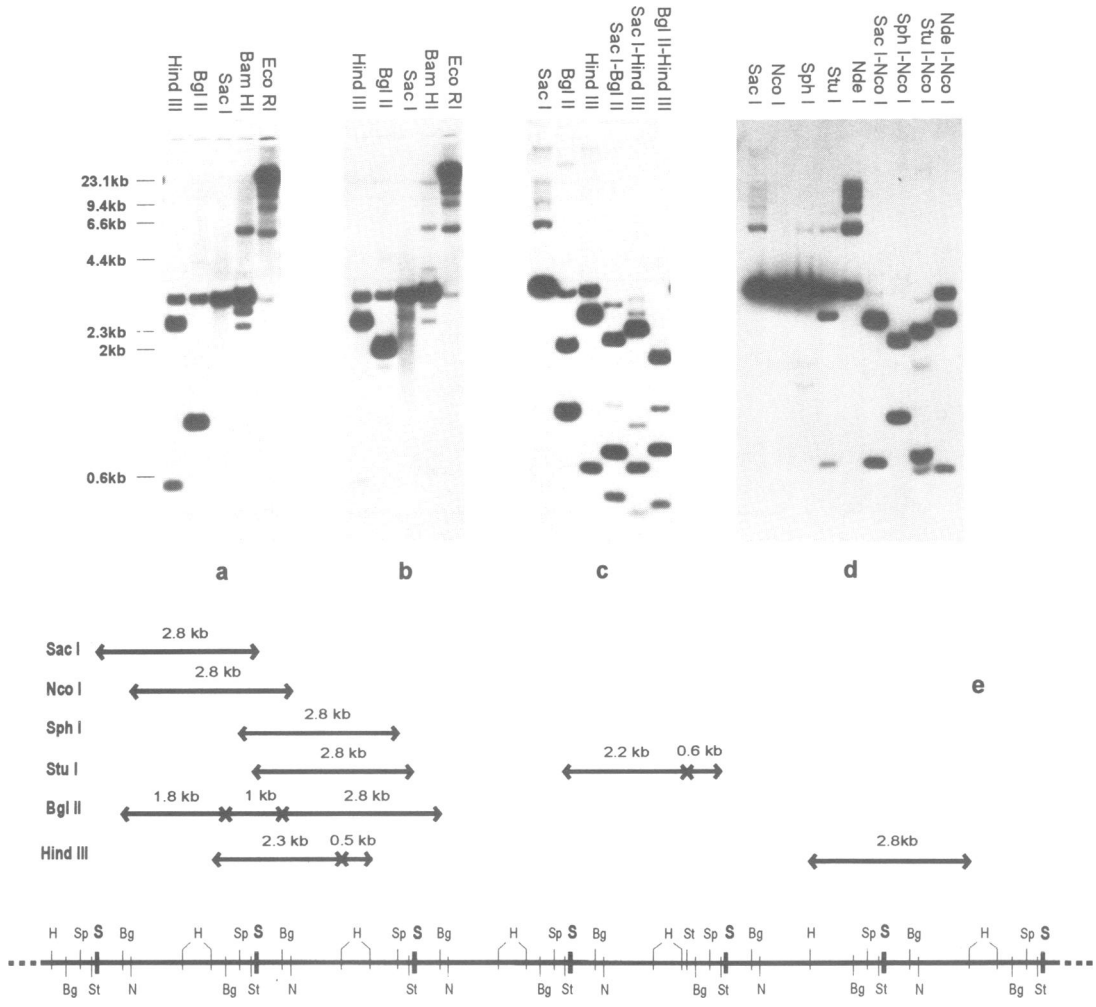

Figure 3. Southern hybridization of CHO-K20 DNA, restricted with the indicated enzymes. Probes are HUCAFF170 (**a**), pHC450 (**b**), pRU2.8b (**c**, **d**). (**e**) Short range genomic organization of HC2sat.

Restriction enzymes having recognition sites in central parts of type '**a**' or type '**b**' units detected significant heterogeneity of genomic repeats. Distribution of *Eco*RV sites defined three major repeat variants (Fig. 4a, b). (i) There were repeat units from which *Eco*RV sites were absent. The majority of these units were clustered and stayed in the unresolved region of the gel. (ii) Repeat units with unique *Eco*RV sites gave a band at 2.8 kb. The positions of *Eco*RV sites in these units were the same as in types '**a**' and '**b**' repeat units (at ~1.6 kb). (iii) The 2.3 kb *Eco*RV band corresponded to repeat variants with two *Eco*RV recognition sites. Double digests indicated that the additional *Eco*RV sites located at positions ~1.1 kb. The apparently absent ~0.5 kb long *Eco*RV fragments of these units could only be revealed by hybridization with pHC450 at low stringency condition (data not shown).

Distribution of *Xba*I, *Bcl*I, and *Dra*I sites were also examined (Fig. 4c). About 70% of repeat units contained *Xba*I sites similarly to the type '**a**' unit. In the majority of these units *Bcl*I and *Dra*I sites were absent. Results of double digests indicated that in the remaining ~30% of repeat units *Bcl*I and *Dra*I recognition sites were linked, and their positions were similar to that observed in the type '**b**' unit (Fig. 4d).

We also investigated the long range organization of the HC2 satellite in CHO cells by pulsed field gel electrophoresis. We found that enzymes containing CpG motifs in their recognition sites (*Mlu*I, *Sac*II, *Eag*I, *Sal*I, *Xho*I, *Cla*I, *Fsp*I) gave smeared and inconsistent fragment patterns. Enzymes with 8 bp recognition sequences (*Asc*I, *Not*I, *Sfi*I) produced fragments >1 Mb, which stayed in the unresolved region of the gel. Figure 5 shows hybridization patterns of CHO DNA cleaved with several enzymes with a 6 bp recognition sequence.

*Ase*I, *Afl*II, *Spe*I, *Ssp*I and *Nsi*I released three bands between 375–440 kb. The rest of the homologous sequences were located on fragments up to 250 kb (Fig. 5a). *Eco*RI, *Nhe*I and *Pvu*II resulted in hybridization patterns similar to that discussed above except that one of the bands of the 375–440 kb size range was absent. Double digests did not alter significantly fragment patterns (Fig. 5b), which indicated that the majority of cleavage sites for

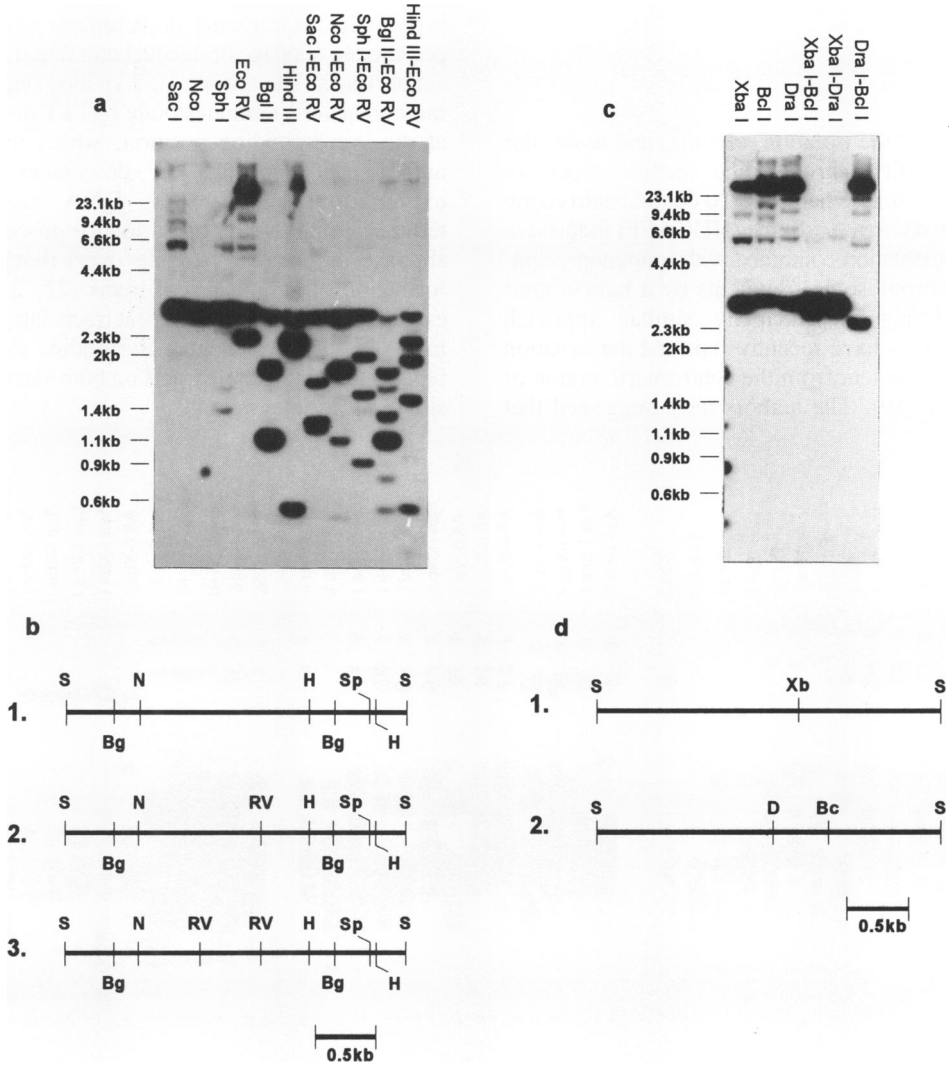

Figure 4. Analyses of repeat unit variants. Southern blots were hybridized with ^{32}P-labeled pRU2.8b fragment (**a**, **c**). Structures of *Eco*RV (**b**) and *Xba*I (**d**) repeat variants. See details in the text.

enzymes used were not farther from each other than several kilobasepairs. On the other hand it is interesting to note that *Eco*RI caused the smallest band of the 375–440 kb range released either by *Ase*I, *Afl*II, *Spe*I, *Ssp*I or *Nsi*I (Fig. 5c lanes 1–5, see later) to disappear.

As mentioned above, *Xba*I, *Bcl*I, *Dra*I and *Eco*RV produced 2.8 kb fragments corresponding to monomer units. The varying amounts of homologous sequences remaining in the unresolved region of the conventional gel were resolved into fragments by pulsed field gel electrophoresis (Fig. 5a, lanes 9–12). *Bcl*I, *Dra*I and *Eco*RV gave fragments up to 220 kb. *Xba*I released an intense 370 kb band and several smaller fragments up to 90 kb. Double digests (Fig. 5c, lanes 6–10) revealed that the 370 kb *Xba*I band corresponded to the smallest band of the 375–440 kb size range produced either by *Ase*I, *Afl*II, *Spe*I, *Ssp*I or *Nsi*I.

*Eco*RI, *Nhe*I and *Pvu*II cleaved the 370 kb *Xba*I fragments (Fig. 5c, lanes 11–13), which was consistent with the absence of the smallest band of the 375–440 kb range from single and double digests of these enzymes. Additional bands appeared above 90 kb in *Afl*II–*Xba*I, *Ase*I–*Xba*I and *Ssp*I–*Xba*I double digests indicated that the 370 kb *Xba*I band represented at least two similar sized fragments.

DISCUSSION

In this paper we described the isolation, cloning, and molecular characterization of the first chromosome specific repetitive centromeric DNA from Chinese hamster. To obtain centromeric DNA, we applied hydroxyurea–caffeine (HUCAFF) induced *in vivo* chromosome fragmentation combined with immunoprecipitation of centromeric chromosome fragments by a human anti-centromere serum. Using a principally similar approach Ouspensky and co-workers have recently reported the isolation of a single copy DNA fragment from the centromeric region of hamster chromosome 1 (19). The authors have suggested that their method has the potential for selective enrichment of centromeric chromatin fractions. Considering the small number of clones examined so far (3 in ref.19 and 8 in this paper), we think that the feasibility of the applied procedure requires further investigation, and can be established only after statistical analysis of a sufficient number of clones.

We demonstrated by restriction mapping of a phage clone (HC90) that DNA regions homologous to the 536 bp HUCAFF170 clone were part of 2.8 kb long repeat units of a novel satellite-like repetitive sequence family specific to Chinese hamster chromosome 2 (HC2sat).

This 2.8 kb repeat element is one of the longest satellite units reported from mammals to date. The longest mammalian satellite described is the 3.8 kb long 1,709 bovine repeat, but the published sequence contains additional interspersed repeat elements embedded in the unit (24). In the sequenced 2.8 kb units of HC2sat neither known interspersed repeats nor extended internal periodicity have been found. Repeat units contained clusters of simple repeats. The type '**a**' unit contained an $(ATTT)_9$ tract. In the equivalent region of the type '**b**' unit, ATTT repeats were replaced by ATTG motifs. In principle, there are alternative explanations for the origin of these simple repeat tracts. (i) It is possible that tetranucleotide repeats were amplified locally by polymerase slippage or unequal crossing over between misaligned simple tracts. The ancestral 2.8 kb long single copy DNA segment might contain a tetranucleotide (ATTT or ATTG or other) motif at the corresponding position, which then could mutate and amplify independently in descendant repeat units. This explanation implies that DNA segments surrounding tetranucleotide motifs could induce processes (e.g. polymerase slippage or unequal crossing over) thought to be involved in formation of simple repeat tracts (22, 23). (ii) An alternative explanation is that simple repeat tracts integrated into repeat units in a site specific manner. Pyrimidine rich degenerated direct repeats flanking simple tracts on both sides may be relics of such site specific integrations.

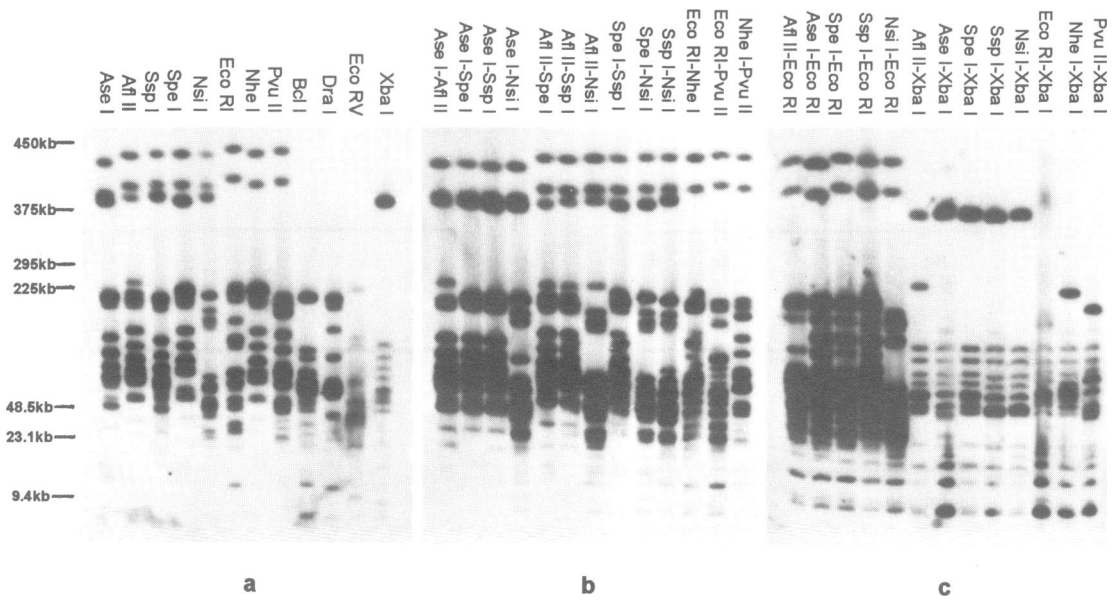

Figure 5. PFGE analyses of the long range organization of HC2sat in the CHO-K20 cell line. The hybridization probe was ^{32}P-labeled pRU2.8b fragment.

Repeat units also contained long $(CA)_n$ tracts. $(CA)_n$ repeats have some interesting features. It has been shown that alternating purine–pyrimidine tracts can adopt a left handed Z-DNA conformation and bind specific nuclear proteins. Studies have also indicated that $(CA)_n$ repeats may be involved in gene regulation and associated with recombination hotspots (24). It is possible, as suggested for other satellite sequences (25) that acquisition of a highly recombinogenic $(CA)_n$ tract was the key event for amplification of the ancestral 2.8 kb sequence.

We confirmed the tandem organization of the novel repetitive sequence by Southern hybridization on restriction enzyme digested genomic hamster DNA. The majority of repeat units had restriction patterns similar to the type 'a' unit present in HC90 phage. Significant heterogeneity could be observed with restriction enzymes having recognition sites in the central parts of repeat units. This observation is consistent with the presence of sequence peculiarities in the central regions of the repeat units. High A–T content, long $d(A-T)_n$ stretches and simple repeat tracts are generally associated with genomic instabilities. For example, novel joints of amplified structures often coincide with A–T rich sequences (26). The low melting temperature of A–T rich segments might facilitate opening of the DNA duplex, which is indispensable for recombination and replication. Long $d(A-T)_n$ runs also have strong influences on chromatin structure by inducing curvature of the axis of the DNA molecule (27). Concentration of such sequence elements in central regions of repeat units might accelerate processes of genomic turnover. We cannot rule out, however, that the homogeneity of 5′ and 3′ ends of repeat units relative to central regions may also reflect some functional significance.

Studies on the long range organization revealed some interesting properties of the HC2 satellite. The majority of enzymes used for digestion of high molecular weight hamster DNA gave two or three large bands between 375–440 kb and numerous smaller bands up to 250 kb. Double digests indicated that the bands between 375–440 kb might represent large homogeneous domains of the tandem array. The homogeneity of domains can be explained if one supposed that they were generated by rapid and local amplification of short segments of the tandem array. Our results also indicated that the domains differed from each other and they might be results of independent short range amplification events. The secondarily amplified DNA segments were not shorter than a repeat unit, because the 2.8 kb long periodicity was general in the tandem array. On the other hand, the homogeneity of domains makes it unlikely that the units of amplifications were significantly larger than several repeat units. Besides this short range amplifications, there were indications for long range amplification events. The 370 kb *Xba*I band might represent at least two similar sized DNA fragments which were generated by duplication of a 370 kb long domain.

In conclusion, our data suggest that the amplification events which led to the formation of the chromosome 2 satellite array took place on different scales and involved different mechanisms. The primary event was the amplification of an ancestral 2.8 kb DNA segment, probably induced by acquisition of a recombination prone $(CA)_n$ tract. This amplification was followed by divergence of the tandem array. The accumulation of mutations was partly counterbalanced by homogenizing processes (unequal crossing over, gene conversion) (28). In the next stage, certain repeat units in the array began to amplify locally. As a result, several hundred kilobasepairs long homogeneous domains formed. The mechanism of these secondary amplifications probably was different from the primary amplification. During the primary amplification recurrent unequal exchanges led to the expansion or contraction of the tandem array. Spreading of repeat variants by this mechanism required several generations. In contrast, secondary amplifications might have occurred via 'rolling circle-like' (29) amplification intermediates. This type of amplification could result in formation of a large number of identical repeat units during a single generation. In parallel, large segments of tandem array were also duplicated by unequal exchanges between misaligned arrays (see 370 kb *Xba*I fragments).

The role of satellite sequences of higher eukaryotes is still the subject of debate. Although, it is generally accepted that pericentric satellites could affect chromosome segregation, the lack of sequence conservation between satellites of different species make it difficult to attribute a specific function to them. The only centromeric repetitive sequence described in Chinese hamster to date is the TTAGGG repeat. It is suspected that these telomere specific repeats are relics of Robertsonian fusion of ancient acrocentric chromosomes (18). The mechanism by which hamster cells can tolerate large internal blocks of $(TTAGGG)_n$ without affecting the telomere function is unknown. The role of telomeric repeats in the pericentric region, if any, is also obscure but it is conceivable that their function is similar to that of different pericentric satellites found in other species. Satellite sequences are constant components of constitutive heterochromatic regions. The regular periodical repetition of a DNA stretch may provide these chromosome segments with some insulating feature. The insulation of the centromeric region from the euchromatic chromosome arms is perhaps essential for proper centromere function and/or protecting genes from some harmful effect of centromeric chromatin.

There are no large, amplified blocks of telomeric repeats at the pericentric region of hamster chromosomes 1 and 2 (18). The HC2 satellite, which is the major constituent of pericentric region of chromosome 2, may be functionally equivalent to centromeric TTAGGG repeats. Some general features of satellites such as concentration of A–T residues, sequence directed curvature and occurrence of simple repeat tracts (27) make it possible that otherwise unrelated sequences can adopt similar chromatin conformation and fulfil similar functions.

Chromosome specific satellites perhaps contribute to the establishment of the order of the interphase nucleus by marking different chromosomes and may play a role in meiotic pairing of homologous chromosomes (30). By affecting these processes, fixation of novel satellites may represent one of the initial steps of speciation.

ACKNOWLEDGEMENTS

We are grateful to Dr A.Udvardy for helpful comments on the manuscript, Drs G.Wang and G.Larson for their help in automatic sequencing. This work was supported in part by grants from the Hungarian National Basic Research Fund (OTKA 926/91), Hungarian National Committee for Technological Development (OMFB 13017/89 and 02177/91) and J nos B styai Holczer Fund.

REFERENCES

1. Singer, M. F. (1982) *Int. Rev. Cytol.*, 67, 67–112.
2. Boyle, A. N., Ballard, S. G. and Ward, D. C. (1990) *Proc. Natl. Acad. Sci. USA* 87, 7757–7761.

3. Okada, N. (1991) *Curr. Opin. Gen. Dev.* 1, 498–504.
4. Martin, S. L. (1991) *Curr. Opin. Gen. Dev.* 1, 505–508.
5. Howard, B. H. and Sakamoto, K. (1990) *The New Biologist* 2, 759–770.
6. Farr, C., Fantes, J., Goodfellow, P. and Cooke, H. (1991) *Proc. Natl. Acad. Sci. USA* 88, 7006–7010.
7. Tyler-Smith, C. and Willard, H. F. (1993) *Curr. Opin. Gen. Dev.* 3, 390–397.
8. Hadlaczky, Gy., Praznovszky, T., Cserp n, I., Keres, J., Péterfy, M., Kelemen, I., Atalay, E., Szeles, A., Szelei, J., Tubak, V. and Burg, K., (1991) *Proc. Natl. Acad. Sci. USA* 88, 8106–8110.
9. Haaf, T., Warburton, P. E., Willard, H. F. (1992) *Cell* 70, 681–696.
10. Gardy, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J., Moyzis, R. T., (1992) *Proc. Natl. Acad. Sci. USA* 89, 1695–1699.
11. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989a) *J. Cell. Biol.* 109, 1963–1973.
12. Masumoto, H., Sugimoto, K. and Okazaki, T. (1989b) *Exp. Cell. Res.* 181, 181–196.
13. Earle, E., Voullaire, L. E., Hill, L., Slater, H. R. and Choo, K. H. A. (1992) *Cytogenet. Cell Genet.* 61, 78–80.
14. Voullaire, L. E., Slater, H. R., Petrovic,V. and Choo, K. A. H. (1993) *Am. J. Hum. Genet.* 52, 1153–1163.
15. Willard, H. F. and Waye, J. S. (1987) *Trends Genet.* 3, 192–198.
16. Boyle, A. L. and Ward, D. C. (1992) *Genomics* 12, 517–525.
17. Nasir, J., Fisher, E. M. C., Brockdorff, N., Disteche, C. M., Lyon, M. F. and Brown, S. D. M. (1990) *Proc. Natl. Acad. Sci. USA* 87, 399–403.
18. Meyne, J., Baker, R. J.,Hobart, H. H., Hsu, T. C.,Ryder, O. A., Ward, O. G., Wiley, J. E., Wurster-Hill, D. H., Yates, T. L. and Moyzis, R. K. (1990) *Chromosoma* 99, 3–10.
19. Ouspenski, I. I. and Brinkley, B. R. (1993) *J. Cell Sci.* 105, 359–367.
20. Brinkley, B. R., Zinkowski, R. P., Mollon, W. L., Davis, F. M., Pisegna, M. A., Pershouse, M. and Rao, P. N. (1988) *Nature* 336, 251–254.
21. Hadlaczky, Gy., Praznovszky, T., Rasko, I. and Keres, J. (1989) *Chromosoma* 97, 282–288.
22. Smith, G. P. (1976) *Science* 191, 528–535.
23. Strand, M., Prolla, T. A., Liskay, R. M. and Petes, T. D. (1993) *Nature* 356, 274–276.
24. Hamada, H.,Petrino, M. G., Kakunaga, T., Seidman, M. and Stollar, B. D. (1984) *Mol. Cell. Biol.* 4, 2610–2621.
25. Skowronski, J., Pluciennicczak, A., Bednarek, A.and Jaworski, J. (1984) *J. Mol. Biol.* 177, 399–416.
26. Hyrien, O., Debatisse, M., Buttin, G. and de Saint Vincent, B. R. (1987) *EMBO J.* 6, 2401–2408.
27. Martinez-Balbas, A., Rodriguez-Campos, A., Garcia-Ramirez, M., Sainz, J., Carrera, P.,Aymami, J. and Azorin, F (1990) *Biochemistry* 29, 2342–2348.
28. Dover, G. (1982) *Nature* 299, 111–117.
29. Delidakis, C., Swimmer, C. and Kafatos, F. C. (1989) *Curr. Opin. Cell Biol.* 1, 488–496.
30. Bostock, C. (1980) *TIBS* 117–119.
31. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
32. Altschul, S. F., Gish, W., Miller, W.,Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410.