

Initiation of DNA replication at the Chinese hamster origin *oriGNAI3* relies on local sequences and/or chromatin structures, but not on transcription of the nearby *GNAI3* gene

Franck Toledo*, Anne-Marie Lachagès, Véronique Mayau and Michelle Debatisse

Unité de Génétique Somatique (URA CNRS 1960), Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Received January 12, 1999; Revised and Accepted February 14, 1999

ABSTRACT

We recently identified a region of preferential replication initiation, *oriGNAI3*, near the 3' end of the Chinese hamster *GNAI3* gene. *oriGNAI3* is co-amplified in mutants selected for *AMPD2* amplification, a process generating chromosomal rearrangements. In this report we have taken advantage of cell lines with truncated and translocated amplified units to show that these rearrangements do not alter the function of *oriGNAI3*. These results indicate that replication initiation at this locus relies essentially on local features. Interestingly, the study of one line in which a rearrangement has disrupted the *GNAI3* gene shows that ongoing transcription of this gene is not required for initiation at *oriGNAI3*. In order to obtain further insight into the sequences and/or chromatin structures required for *oriGNAI3* function, we have analyzed the DNase I sensitivity and nucleotide sequence of the region. The features important for replication initiation appear to cluster in a 7–12 kb region which includes *oriGNAI3*.

INTRODUCTION

In mammalian cells, DNA replication has been shown to initiate at defined chromosomal loci, termed initiation regions, but these loci are still poorly characterized (1,2). As a consequence, although mammalian homologs to the yeast replication initiation factors have been identified (3–5), their binding to specific DNA sequences has not been demonstrated, and their role in promoting initiation is not proven so far. This situation results in part from the lack of efficient assays for autonomous plasmid replication in mammalian cells. Certain experiments indicate that plasmids containing sequences previously identified as chromosomal initiation regions cannot replicate autonomously (6), and others suggest that the size rather than the sequence of the cloned DNA is important for autonomous replication in mammalian cells (7). These results have raised doubts as to whether specific sequences are required to initiate DNA replication in mammalian cells, and have also led to the proposal that plasmids lack the structural

chromatin features that may focus initiation to some loci in chromosomes (7,8). For this reason, the search for structures or sequences involved in replication initiation has subsequently been carried out within chromosomal contexts.

The importance of chromosome architecture, rather than specific DNA sequences, in specifying replication initiation to particular loci was further supported when the replication of the Chinese hamster dihydrofolate reductase (*DHFR*) chromosomal locus was studied in *Xenopus* egg extracts (9,10), and more recently in Chinese hamster cells (11). In some respects, the replication pattern of the human β -globin locus also supports this hypothesis: in Hispanic thalassemia, the deletion of 35 kb of sequence including the locus control region leads to an alteration in the chromatin structure, gene expression and replication timing of the region. This deletion was also found to abrogate replication initiation at an origin located 50 kb downstream, within or close to the β -globin gene (12). On the other hand, the importance of more local chromatin structures or sequences for replication initiation was also demonstrated for this origin in Lepore thalassemia, for which a deletion of sequences at the replication origin also abolished initiation (13). Although these studies provided important information concerning the β -globin origin, an approach relying on deletions occurring in thalassemic patients could not be used to study other initiation loci. A more general strategy involves transfection of cells with a construct containing a putative replication origin, to determine whether or not the origin can function when integrated at one or several ectopic locations (8). This approach, which can be refined with targeted deletions using site-specific recombinases (14), enables the study of the role of local sequences or structures in replication initiation. However, long range chromatin structures or sequences important for origin function *in loco* may escape detection when the origin is studied at ectopic locations, as the chromatin environment at the integration sites might compensate their absence. For example, the sequences deleted in Hispanic thalassemia are required for initiation at the β -globin origin *in loco* (12), but not at ectopic locations (14). Likewise, sequences or chromatin structure at the promoter of the *DHFR* gene appear to be required *in loco* for efficient replication initiation downstream of this gene (15), but a 17 kb insert that does not contain the *DHFR* promoter

*To whom correspondence should be addressed. Tel +33 1 45 68 85 71; Fax: +33 1 40 61 31 71; Email: ftoledo@pasteur.fr

can sustain initiation ectopically (8). DNA replication initiation in mammalian cells is, therefore, still poorly understood as mammalian replication origins must be studied in a chromosomal context, preferably *in loco*. A spectrum of different replication origins has to be studied, with a variety of strategies, in order to get further insight into the mammalian initiation process.

Recently, we reported the identification of a replication initiation region, termed ori*GNAI3*, in the intergenic region between the Chinese hamster *GNAI3* and *GNAT2* genes (16), two genes encoding respectively the α 3 and α 2 subunits of GTP-binding proteins (17,18). The properties of this origin strengthened the notion that specific sites of replication initiation exist in mammalian genomes, since ori*GNAI3* was detected as a narrow region of preferential initiation by both competitive PCR and two-dimensional gel electrophoresis (16). Therefore, ori*GNAI3* appeared to represent an attractive model for the study of sequences or chromatin structures governing initiation. With this aim in mind, we took advantage of the co-amplification of the ori*GNAI3* region with the nearby adenylate deaminase 2 (*AMPD2*) gene, in mutant cells which had been selected for *AMPD2* gene amplification (19). In such mutants, amplified units are several hundred kilobases long and often contain the *GNAI3*, *GNAT2*, *AMPD2* and *GSTM* genes (*GSTM* encodes a glutathione S-transferase of the μ family), as well as an uncharacterized gene that maps between *AMPD2* and *GSTM* (20,21). Amplification was useful here because major chromosomal rearrangements are generated during this process (22,23), and cell lines with amplification-associated chromosomal rearrangements that have altered sequences flanking the ori*GNAI3* region have previously been described (19). We show here that in some lines all amplified sequences are rearranged, and most of them have been translocated to other chromosomes. Importantly, amplified units were translocated as clusters that are several megabases, or tens of megabases, in length. This property is expected to reduce the potential position effects near the translocation sites, by contrast with the experiments mentioned above which involved phage or cosmid constructs integrated at ectopic locations (8,14). Using quantitative PCR we demonstrate that ori*GNAI3* function is maintained in cells of these lines. This shows that initiation of DNA replication at ori*GNAI3* is governed by local specific chromatin structure and/or sequence features. The analysis of one line in which a rearrangement disrupts the *GNAI3* gene shows that ongoing transcription of this gene is not essential for replication initiation. The requirements for initiation at ori*GNAI3* were further analyzed by studying the sensitivity to DNase I and the nucleotide sequence of the region. Both analyses strengthen the notion that local chromatin structures and/or sequences are important for the function of ori*GNAI3*, and suggest that such features cluster in a 7–12 kb region.

MATERIALS AND METHODS

Cell lines and culture conditions

Lines 474 and 633 are coformycin-resistant mutants isolated after three steps of selection from GMA32, a Chinese hamster lung fibroblastic cell line (19). Both were grown in Eagle medium supplemented with adenine, azaserine, uridine and 25 μ g/ml of coformycin. Southern and dot blot analyses previously showed that cells of lines 474 and 633 contain 100–150 and copies of the *AMPD2* region (19). The 42 cell line was isolated after two steps of selection from GMA32 cells and grown in a medium containing 5 μ g/ml of coformycin (16,19).

In situ hybridization

Fluorescent *in situ* hybridization (FISH) analysis was performed essentially as described (23). Cosmids 56 A3 and 56 Y1B were previously described (19); cosmid 61 C2 was recovered by chromosome walking from cosmid 61 W 14. Cosmid 56 A3 was labeled with digoxigenin by random priming and cosmids 61 C2 and 56 Y1B were biotinylated by nick translation. The probes were recovered by filtration through Quick Spin Columns (Sephadex G-50; Boehringer). Slides, treated with RNase and Proteinase K as described (24), were hybridized with equal mixtures of cosmids 56 A3 and 61 C2, or 56 A3 and 56 Y1B. The hybridization mixture, consisting of 10 μ l total volume per slide containing 50 ng of each probe, 10 μ g of sonicated GMA32 genomic DNA, 50% formamide, 1% Tween 20 and 10% dextran sulfate in 2 \times SSC, was heated for 10 min at 70°C, placed at 37°C for 20 min, then applied to slides pre-heated to 70°C for 2.5 min in 70% formamide, 2 \times SSC, pH 7.0. Hybridization was at 37°C overnight. Biotin was revealed with alternating layers of fluoresceinated avidin and biotin-conjugated goat anti-avidin antibody (Vector Labs); digoxigenin by successive treatments with mouse anti-digoxin, rhodamine-conjugated anti-mouse and anti-rabbit antibodies (Sigma). Slides were mounted with Vectashield (Vector labs), an antifading preparation containing diaminidino-phenylindol.

Competitive PCR analysis

DNA extraction and size-fractionation by sedimentation through neutral sucrose gradients was performed as previously described (16). In a first series of experiments performed in the dark, 2 \times 10⁷ exponentially growing cells of line 474 were labeled for 4 min with 10⁻² M BrdUrd. The labeling medium was removed, cells were washed and scraped, then nuclei were recovered as described (25). Pellets were resuspended in 1 ml of 10 mM Tris pH 8.0, 50 mM EDTA, then DNA preparations were denatured by addition of 1 ml of 1 M NaCl, 0.5 M NaOH, 50 mM EDTA, 0.4% Sarkosyl, followed by incubation at 45°C for 30 min. DNA strands were then separated according to size by sedimentation through a sucrose gradient as described (16). Fractions containing 1–2 kb strands were pooled, purified on anti-BrdUrd immunaffinity chromatography columns, and eluates were quantified by competitive PCR as described (26). In the following series of experiments, DNA was prepared according to the nascent DNA extrusion method (27), with modifications. 10⁷ exponentially growing cells of lines 474 or 633 were washed, scraped, and nuclei were recovered as above. Pellets were then washed three times with 10 mM Tris pH 8.0, 10 mM EDTA, resuspended in this buffer, treated with 25 μ g of RNase at 37°C for 15 min, then at 55°C overnight with 8 mg of Proteinase K in 10 mM Tris pH 8.0, 10 mM EDTA and 0.33% SDS. DNA strands were then extracted with an equal volume of 1:1 phenol/chloroform, precipitated with ethanol, resuspended in TE buffer (10 mM Tris, 1 mM EDTA), then separated according to size by sedimentation through a sucrose gradient as described (16).

The primers and competitors used for the quantitative PCR reactions at five of the seven analyzed loci have been described (16). For each additional locus analyzed here, four primers had to be designed, as described (16): two were chosen in order to amplify DNA fragments of 200–300 bp (*ef*, external forward and *er*, external reverse); the two others (*if*, internal forward and *ir*, internal reverse) consisted of two complementary 5' tails of 20 nt unrelated to genomic sequences (tail1: 5'-GTCGACGGATC-

CGAATTCGT-3'; tail2: 5'-ACGAATTCGGATCCGTCGAC-3'), linked to sequences on the 3' end complementary to genomic targets, allowing competitors to differ from targets by an addition of 20 nt. Sequences (5'→3') of these primers are: primer set γ : *ef*, TCTGTTCACTGGTATGTTCC; *er*, CTTACGTTCTTGAGG-TG; *if*, tail1+GAGGGCTGTGCTGGGTAGAAA; *ir*, tail2+GCA-CAGATTCCAGCAGGGA. Primer set δ : *ef*, TGGTTCACA-GACATAATCAGGC; *er*, GGATGAGAGAGGCTATGGATC; *if*, tail1+CCCTGGACTGAAGCAAATGG; *ir*, tail2+AAGAGGA-CTACAGGCACGCA.

The localizations of the amplification products obtained with these external primers are: γ , 7576–7802 and δ , 9850–10087 (see below for sequence details). Quantification of each genomic target sequence was performed as described, by coamplifying known and increasing amounts of the corresponding competitor with a fixed amount of a nascent strand preparation (16).

DNase I hypersensitivity mapping

The protocol used was adapted from RecillasTarga *et al.* (28). All operations were carried out at 4°C, starting from four 23 cm-square dishes each containing 3.5×10^7 cells of line 42. The culture medium was removed and each dish was washed with 30 ml of ice-cold KT-NP-40-PMSF (50 mM KCl, 50 mM Tris-HCl pH 7.4, 10 mM EDTA with 1% NP-40 and 100 nM PMSF), cells were scraped, layered over 20 ml of cold 20% glycerol in KT, centrifuged at 4°C for 10 min at 2000 r.p.m., washed with 20 ml of HKP-EDTA (5 mM HEPES pH 7, 85 mM KCl, 0.5 mM spermidine, 0.25 mM PMSF, with 10 mM EDTA), centrifuged as before and resuspended in 18 ml of HKP, to which 1.25 mM CaCl₂ was added. The resulting solution was aliquoted into six tubes and DNA was digested with 0–1400 U of DNase I for 20 min at 4°C. The reaction was stopped by addition to each tube of 1 ml of lauryl sarcosine 1%, 400 mM EDTA. One milligram of Proteinase K was added to each tube, and incubation was carried out overnight at 42°C. DNA was then extracted, resuspended in TE and stored at 4°C. Aliquots of these preparations were digested with the chosen restriction enzymes and restriction fragments were separated by agarose gel electrophoresis, blotted and hybridized with appropriate probes.

Sequence analysis

The analyzed sequence (19 400 bp) results from linking the sequences for *GNAI3* (17,29), the intergenic *GNAI3*–*GNAT2* region (16) and *GNAT2* (18). DNA unwinding analysis was performed with the Thermodyn program. DUEs were searched for as described (30): a single major DUE was identified as 400 contiguous nucleotides with an average ΔG inferior to the mean ΔG of the entire sequence by at least 10 kcal/mol. DNA bending was analyzed with GentlBen at window width of 120 bp, to calculate ENDS ratios. Fitzgerald *et al.* have shown that an ENDS ratio value >1.15 is a valid criterion for bending prediction (31). Therefore only such values were reported here. Search for the initiation consensus sequence (ICS), RIP60 binding sites and CpGs clusters, were essentially performed with GCG (v.8, Wisconsin Package) or DNASIS (Hitachi). The ICS is defined as WAWTTD-DWWWHDHWGWHMAWTT. A 20/21 match to this sequence was found in six eukaryotic replication initiation regions (30), and searched for here. RIP60 binding sites are proposed to correspond to one or several (TATTAT) tracts (32). Clusters of CpG dinucleotides were searched for at a window width of 40 bp (33).

RESULTS

Rationale of the study

ori*GNAI3* was first identified in line 42, an amplified mutant cell line (16) which was chosen because it retains many of the properties of GMA32, the unamplified line from which it derives. In cells of line 42, amplified units map on the long arm of one chromosome 1, like the normal copies in unamplified cells. Furthermore, at least for the cloned 150 kb of sequence surrounding the *AMPD2* gene, the restriction maps are identical in amplified units of line 42 and in unamplified cells (16,19). We have also shown previously that the expression patterns of the genes within the cloned region are similar in amplified and unamplified cells. In both lines the *GNAI3*, *AMPD2* and *GSTM* genes are expressed, but not *GNAT2*, which is a retina-specific gene (20). In the present study, we have analyzed the consequences on ori*GNAI3* function of chromosomal rearrangements occurring close to the origin. Gene amplification is a process that generates chromosomal rearrangements (22,23), allowing the recovery of cell lines containing rearranged amplified units (19). To be suitable for the present study, such cell lines had to contain only rearranged amplified units, so that the identified replication patterns could unambiguously be attributed to the rearranged copies. As detailed below, two cell lines, designated 474 and 633, fulfilled this requirement.

Replication initiation in the *GNAI3*–*GNAT2* region in cells of line 474

Previous Southern blot analyses have shown that amplified units in line 474 are characterized by a rearrangement that has removed the 5' part of the *GNAI3* gene, thereby abolishing its transcription, as indicated by northern blot analyses (34,35). In the present study we characterized cell line 474 at the cytological level, by FISH using cosmids 61 C2, 56 A3 and 56 Y1B as probes (Fig. 1A). Results show that all amplified units in cells of line 474 are rearranged: they lack the sequences homologous to cosmid 61 C2, since a single signal for this cosmid is detected on an unamplified chromosome 1, at its normal location. Furthermore, rearranged amplified units are found on two chromosomes: a few map on the long arm of one chromosome 1 homolog, but most are located on a small, unidentified chromosome (Fig. 1A). Thus cytological results indicate two differences between lines 474 and 42: in cells of line 474 most amplified units have been translocated to a marker chromosome, and all amplified units have lost the 5' part of the *GNAI3* gene.

ori*GNAI3* was identified by quantifying short nascent strands at five different loci distributed along 15.5 kb of the region, using competitive PCR with primer sets α to ϵ (16). As illustrated in Figure 1B, molecular analyses have indicated that the rearrangement in line 474 occurred ~1 kb on the left of primer set α (35), leaving the previously analyzed 15.5 kb region unaltered. This rearrangement led to an inverted organization of the amplified units, so that the amplified structure comprises only sequences located downstream of the recombination site (35). A first series of experiments was performed according to the original PCR replicon mapping technique (36), which relies on the quantification of purified BrdUrd-labeled short nascent strands. Exponentially growing cells of line 474 were labeled with a short pulse of BrdUrd and 1–2 kb BrdUrd-labeled molecules were recovered (see Materials and Methods). Quantification of these molecules was achieved at

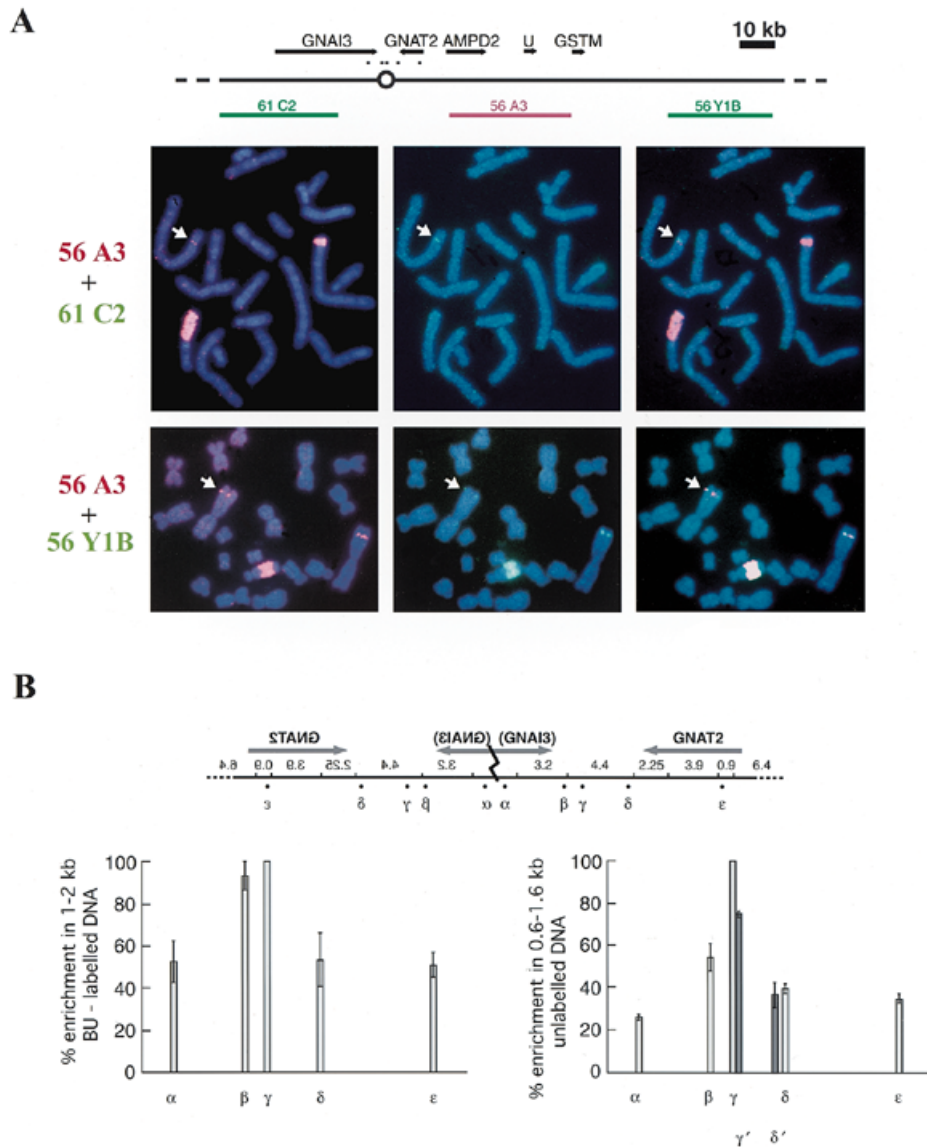


Figure 1. Analysis of *oriGNAI3* in cells of line 474. (A) Two-color FISH analysis of the cell line. Top row: map of the unrearranged region, with the location of the cosmid probes used in this study (61 C2, 56 A3, 56 Y1B), *oriGNAI3* (open circle), primer sets (dots) and genes (arrows indicate transcription direction, U marks the gene with unknown function). Middle row: a typical cell hybridized with cosmid probes 56 A3 (red) and 61 C2 (yellow). Photographs from left to right show, respectively, signals with the 56 A3 probe only, the 61 C2 probe only, or both probes. Arrows indicate the unrearranged single copy on the unamplified chromosome 1q. Bottom row: a cell hybridized with cosmid probes 56 A3 (red) and 56 Y1B (yellow). Results are presented as above. (B) Competitive PCR analyses of *oriGNAI3* in line 474. Top: map of the rearranged amplified units. The amplification-associated rearrangement (shown by a zigzag line) has created a palindromic junction between the two 3' ends of the *GNAI3* gene (35). Gray arrows: directions (5'→3') of the genes; vertical bars, *EcoRI* restriction sites; numbers, size (in kb) of *EcoRI* restriction fragments; dots, locations of primer sets α to ε (see text for details). Below, a summary of the PCR quantification analyses is presented. Left: quantification of 1–2 kb BrdUrd-labeled molecules. Results are the average of three quantification experiments from two independent BrdUrd (BU) DNA preparations, plotted against the map of the region. In order to eliminate fluctuations resulting from variable efficiencies in the recovery of BrdUrd-labeled molecules, the amount of molecules found with each primer set was expressed as a percentage of the quantity of molecules determined with primer set γ. Bars indicate confidence limits at a 0.95 significance level. Right: a similar representation of the quantification of 0.6–1.6 kb unlabeled molecules. Results are the average of seven quantification experiments from two DNA preparations. In these experiments two primer sets (γ' and δ', dark gray histograms) were added to the five previously used (light gray histograms).

the five loci (α to ε), by primer competition between the genomic target and a competitor molecule differing from the genomic sequence only by a 20 bp insertion as previously described (16,36). The results of three quantification experiments, carried out with independent DNA preparations, are presented in Figure 1B (left). These results are very similar to those obtained when 1–2 kb long BrdUrd-labeled DNA preparations from line 42 were quantified (16): a peak of molecules is detected with

primer set γ, the only primer set that maps at or close to *oriGNAI3*. Primer set β, which maps ~700 bp away from *oriGNAI3*, allows the detection of nearly as many molecules as primer set γ. Primer sets α, δ and ε disclose comparable amounts of molecules, about half as many than at locus γ.

We have shown previously that the signal-to-noise ratio (i.e. the number of molecules at locus γ divided by the number of molecules at loci α, δ or ε) could be significantly improved by

using a procedure that does not rely on BrdUrd labeling, probably because the *GNAI3*–*GNAT2* region is highly A-T-rich and prone to breakage when labeled by BrdUrd (16). We also found that the number of molecules at locus β could be decreased by quantifying smaller molecules, because most molecules at locus β correspond to initiation events that occurred at or close to primer set γ (16). Therefore, in a second series of experiments, we applied these modifications to the study of cells of line 474. In order to purify nascent strands without labeling cells with BrdUrd, we used the nascent strand extrusion procedure, which has been successfully used in combination with PCR quantification for the analysis of several initiation regions (27), and 0.6–1.6 kb molecules were chosen for quantification. The results of this second series of experiments are reported in Figure 1B (right). They were again very similar to those obtained with cells of line 42 with comparable experimental conditions: the signal-to-noise was improved, and the number of molecules at locus β decreased. In this series of experiments, the analysis was expanded to two additional loci on the right of primer set γ : primer sets γ' and δ' . Results show that the number of molecules at locus δ' is similar to those found at loci α , δ and ϵ , so if initiation events occur at this locus, they are very rare. The number of molecules found at locus γ' remains inferior to the number of molecules at the adjacent locus γ , suggesting that most molecules quantified at γ' correspond to initiation events that occurred at, or on the left of locus γ (similarly to what was observed at locus β when longer molecules were quantified; see above). Thus, the results obtained with the two additional loci suggest that if initiation events occur on the right of locus γ , they are not frequent.

In conclusion, with or without BrdUrd labeling, the replication pattern of the *GNAI3*–*GNAT2* region is similar in cells of lines 474 and 42. These results indicate that all the features essential for replication initiation at *oriGNAI3* are located downstream of the rearrangement site in cells of line 474, and that a lack of *GNAI3* transcription has no effect on *oriGNAI3* function. Furthermore, if *oriGNAI3* was only active in cells of line 474 in the minority of rearranged amplified units that map on chromosome 1, the background generated by the large majority of amplified units on the smaller chromosome would be expected to increase, leading to a poor signal-to-noise ratio. Since similar signal-to-noise ratios were found with competitive PCR in cell lines 42 and 474 regardless of the experimental procedure, we conclude that the translocation of amplified units to a small chromosome in cell line 474 does not alter the function of *oriGNAI3*.

Replication initiation in the *GNAI3*–*GNAT2* region in cells of line 633

In order to determine if distant features on the right of *oriGNAI3* are required for replication initiation, we studied cells from line 633, another mutant line amplified for the *AMPD2* gene (19). Cells from this line were analyzed by FISH using the same cosmid probes as before (Fig. 2A). Results show that amplified sequences are spread on five different chromosomes. Furthermore, all amplified units in cell line 633 are rearranged: they contain sequences homologous to cosmids 61 C2 and 56 A3, but not 56 Y1B (Fig. 2A). In these cells the rearrangement has removed sequences downstream the *GSTM* gene and led to an inverted organization of the repeated units: this rearrangement site lies some 45 kb away on the right of the region previously analyzed by PCR (19 and Fig. 2B). Nascent strands were purified by the

extrusion method, and the replication pattern of the *GNAI3*–*GNAT2* region was analyzed by PCR, using the seven primer sets previously described. Results show that the replication pattern in cells of line 633 is similar to those observed in cells of line 474 (Fig. 2B, to be compared with Fig. 1B, right) and in cells of line 42 (16). Thus the deletion of sequences and chromosome structures downstream the *GSTM* gene does not alter the function of *oriGNAI3*. Although Figure 2B shows that replication initiation occurs at *oriGNAI3* in cells of line 633, the signal-to-noise ratio in these cells is slightly inferior to the one in cells of line 474 (Fig. 1B, right), but we cannot clearly establish that this difference is significant. Therefore, we conclude that in cells of line 633, *oriGNAI3* functions in most, but maybe not all, translocated amplified units.

Identification of DNase I hypersensitive sites in the region between the two rearrangement sites

The function and precise position of *oriGNAI3* is identical in cells from lines 42, 474 and 633, indicating that the rearrangements of amplified units have no effect on the origin and that the function of *oriGNAI3* relies essentially on sequences or chromatin structures that map between the two rearrangement sites. To gain further insight into the chromatin structure of the locus, we searched for DNase I hypersensitive sites, which reflect gaps in the nucleosomal array and may result from the binding of proteins at specific sites (37–39). In the present study, some of these hypersensitive sites might correspond to the binding of proteins important for replication initiation (40,41). DNase I hypersensitive sites were searched for within the 60 kb region that lies between the rearrangement sites (Fig. 3). Isolated nuclei from line 42 were treated with various concentrations of DNase I prior to DNA extraction and digestion with appropriate restriction enzymes (Materials and Methods). This analysis revealed a striking difference between the region encompassing the 3' ends of the *GNAI3* and *GNAT2* genes and the *GNAI3*–*GNAT2* intergenic region (Fig. 3B, region 1), and the rest of the studied domain, spanning the *AMPD2* gene, the unidentified gene (*U*) and the 5' end of the *GSTM* gene (region 2).

In region 2, unmethylated CpG islands have previously been identified at the 5' ends of the *AMPD2* gene, the unidentified gene and the *GSTM* gene (21). DNase I hypersensitive sites were found here only at or close to these CpG islands (Fig. 3). An example of this is shown in Figure 3A: when a 12.5 kb *SacI* restriction fragment (S) spanning the *GNAT2*–*AMPD2* intergenic region and the 5' part of the *AMPD2* gene is hybridized with probe d, three hypersensitive sites are revealed at the level of the CpG island of the *AMPD2* gene only. When the same probe is used to analyze the overlapping 16 kb *EcoRI* restriction fragment (E1), spanning the downstream sequences of the *AMPD2* gene and part of the *AMPD2*–*U* intergenic region, no hypersensitive sites were found. The contiguous 7.3 kb *EcoRI* fragment (E2), contains two hypersensitive sites, at the CpG island of the unknown gene. Similarly, near the rearrangement site of cell line 633, the DNase I hypersensitive sites detected were within the CpG island of *GSTM* (not shown). Thus in region 2, the positions of the DNase I hypersensitive sites strongly suggests that they relate to the proposed functions of CpG islands (42–44), rather than to the function of *oriGNAI3*.

Hypersensitive sites to DNase I were also found in region 1, which includes *oriGNAI3*, but in which no unmethylated CpG island had been identified (21). Five major hypersensitive sites

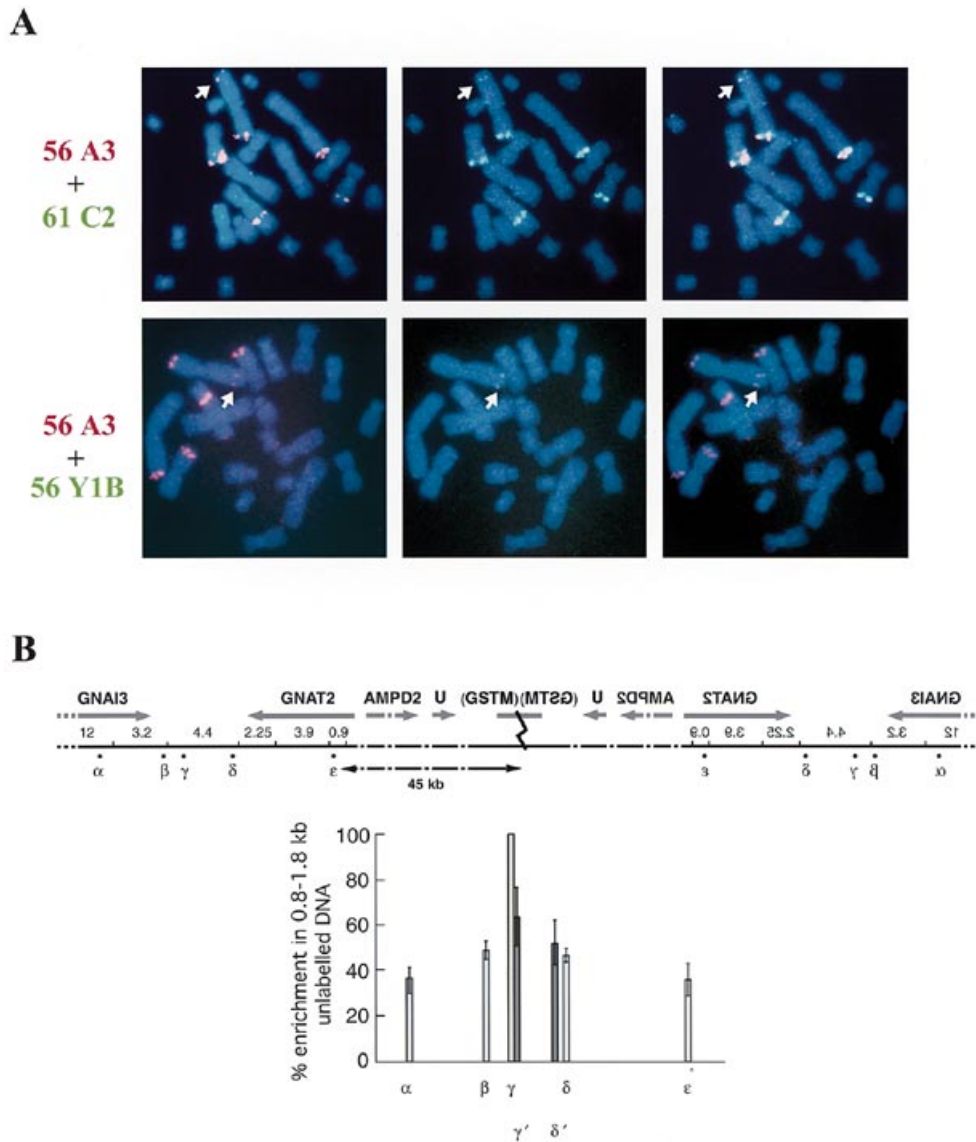


Figure 2. Analysis of *oriGNAI3* in cells of line 633. **(A)** Two-color FISH analysis of the cell line. Results are presented as in Figure 1A. **(B)** Competitive PCR analyses of *oriGNAI3* in line 633. Top: map of the rearranged amplified units, with only the *GNAI3*–*GNAT2* region drawn to scale. Symbols as in Figure 1B. The amplification-associated rearrangement has created a palindromic junction within the *GSTM* gene, 45 kb away from the region analyzed by competitive PCR (19). Below: summary of the PCR quantification analyses. Results are the average of eight quantification experiments carried out on 0.8–1.8 kb unlabeled molecules from two independent DNA preparations. Quantifications are represented as in Figure 1B (right).

were found. As shown in Figure 3, a 12.4 kb *Xba*I restriction fragment (X), detected using probe b, allowed two hypersensitive sites flanking *oriGNAI3* to be mapped: one at the 3' end of the *GNAI3* gene, the other within the intergenic region between *GNAI3* and *GNAT2*. On the left of this fragment, no DNase I hypersensitive site was found, as shown by hybridization of a 7.8 kb *Eco*RV fragment (EV) with probe a. The 12.2 kb *Hind*III fragment (H), which overlaps the *GNAT2* gene and most of the *GNAT2*–*AMPD2* intergenic region, enabled three other major sites to be mapped within *GNAT2*. These three sites map within a region devoid of any transcriptional activity, since *GNAT2* is unexpressed in the fibroblasts analyzed here (20). Thus, the five hypersensitive sites spread across the 12 kb region 1 differ strikingly from the hypersensitive sites in region 2. Although

some of these sites may reflect the binding of factors involved in repressing the transcription of *GNAT2*, it seems likely that others result from the binding of factors involved in replication initiation at *oriGNAI3*. Another particular feature of this 12 kb region is that it contains many additional, weaker sites, as attested from the faint bands observed with both fragments X and H (Fig. 3). This suggests that the whole region, from the 3' end of the *GNAI3* gene to a large part of the *GNAT2* gene, is easily accessible for DNase I, a property that may also relate to the replication initiation process.

Sequence analysis of the *GNAI3*–*GNAT2* region

The analyses of sensitivity to DNase I described above suggest the existence of a particular chromatin structure in region 1

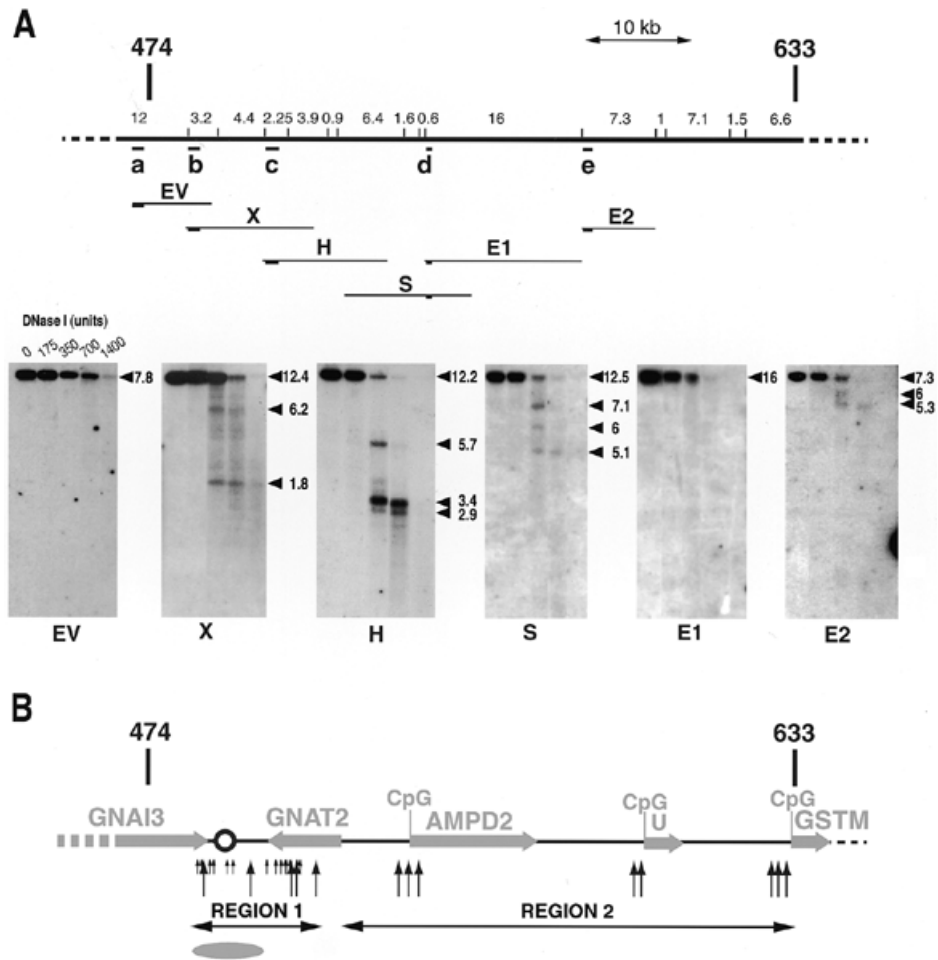


Figure 3. Chromatin structure in the 60 kb region between the two rearrangement sites. (A) DNase I hypersensitive site mapping experiments. Top: *EcoRI* map of the region. Large vertical bars, rearrangement sites in cell lines 474 or 633; small vertical bars, *EcoRI* restriction sites; numbers, size of *EcoRI* restriction fragments. Below the map are shown the probes (black boxes, a to e) and analyzed fragments (EV, X, H, S, E1, E2). The corresponding restriction sites are: EV, *EcoRV*; X, *XbaI*; H, *HindIII*; S, *SacI*; E, *EcoRI*. For each photograph below, lanes from left to right correspond to DNA digested with 0, 175, 350, 700 and 1400 U of DNase I. Arrowheads point to the full-length fragments and, when applicable, to the major fragments produced by DNase I digestion; corresponding sizes are indicated. (B) Summary of the analyses. On the same scale as in (A), this map shows the respective location of genes and hypersensitive sites. Gray arrows, genes; CpG, CpG islands; large and small black arrows, major and minor DNase I hypersensitive sites, respectively. The pattern of DNase I hypersensitivity markedly differs between region 1 and 2 (see text). The gray oval indicates the region in which sequence features proposed to be involved in replication initiation are clustered.

(Fig. 3). We thus analyzed the nucleotide sequence of a 19.4 kb stretch (16–18,29) including this region (Materials and Methods). Although mammalian replication origins are still ill-defined, several sequences or structural features have been proposed to be involved in replication initiation. They can be classified into three categories. (i) Those that play a role in the initiation of replication in yeasts or eukaryotic viruses: they include DNA unwinding elements (DUEs), which are inherently unstable duplex DNA segments (45,46). (ii) Those for which a role in replication initiation has been proposed but remains controversial: among them are regions of attachment to the nuclear matrix (MARs) and stably bent DNA (47–51). (iii) Those for which there is no current evidence of a role in replication initiation, but which are frequently found at mammalian replication origins: these include the ICS, binding sites to RIP60, and clusters of CpG dinucleotides (30,32,33). Most of these candidate motifs are characterized by a rather relaxed sequence conservation and

may be considered as features of local chromatin structure. All motifs considered above were found clustered in a 7 kb region centered over *oriGNAI3* (Fig. 3B). This 7 kb region includes part of the large MAR previously identified in this region (21), one DUE, five motifs of DNA bending, one ICS, three clusters of CpG dinucleotides and eight putative binding sites to RIP60, whereas the remaining 12.4 kb of analyzed sequences contain a single cluster of CpG and none of the other motifs (not shown). Interestingly, the 7 kb region includes the 3' untranslated region (3'UTR) of the *GNAI3* gene (Fig. 3B): this may suggest a partial overlap of the signals important for replication initiation and those specifying transcription termination. Furthermore, it is probably worth noting that the DUE detected here co-maps with one of the major DNase I hypersensitive sites identified above: a similar situation was observed at the human *lamin B2* origin, and was shown to correspond to a footprint proposed to relate to replication initiation (41).

DISCUSSION

In mammalian cells, a role for specific sequences or structures in the initiation of DNA replication was first suggested when a recombinant phage containing a candidate replication origin, transfected into mammalian cells and integrated at random sites, appeared to initiate DNA replication (8). Using such an approach, the consequences of rearrangements within or close to the initiation region could not be studied, however, because of the potential position effects on the rearranged versions of the origin as a result of their different chromosomal locations, a major inconvenience given that chromosome structure is an important factor in mammalian replication initiation (7,9,11). An approach that circumvents this problem was recently applied to the origin at the β -globin locus. It relies on the use of site-specific recombinases to integrate a replication origin at a single ectopic chromosomal site, and then to produce rearranged versions of this origin at the same site (14). While this is a powerful procedure for the analysis of the local features involved in initiation, the origin is still studied at an ectopic location, within a different chromosomal environment, and this has major consequences. Indeed, the sequences deleted in Hispanic thalassemia are required for initiation at the β -globin origin *in loco*, but not at the ectopic site, probably because the plasmid construct is integrated next to a *cis*-acting sequence which exerts the same influence on the origin than the sequences deleted in Hispanic thalassemia (12,14).

In the present study, we have used an alternative strategy, which has the advantage of reducing the effects of ectopic chromosomal environments, to analyze the consequences of chromosomal rearrangements on the function of a replication origin. We have studied replication patterns of the *GNAI3*–*GNAT2* region in the 474 and 633 cell lines, which both contain amplified units translocated to one or several ectopic locations. In this system, the influence of a new chromosomal context is expected to be minimal, since the clusters of amplified units that were translocated are several megabases, or tens of megabases long (as indicated by FISH data; Figs 1A and 2A). Accordingly, the impact of rearrangements close to the origin can be studied, regardless of the location of clusters of amplified units. Furthermore, among the previously described mutant cell lines with rearranged amplified units, lines 474 and 633 were chosen for three reasons: (i) as shown with the present FISH analysis, in these lines all amplified units are rearranged, so that the identified replication patterns can unambiguously be attributed to the rearranged copies (Figs 1A and 2A); (ii) previous molecular analyses had shown that rearranged amplified units in these cell lines are homogeneous in structure (19,35), so that the consequences of a single rearrangement can be studied; and (iii) in these lines the rearrangements have linked the truncated amplified sequences in an inverted orientation (19,35). This situation is more adequate for the present study, since deletions close to the origin occurred, without leading to the junction of unrelated sequences that could compensate for those deleted.

We have used the competitive PCR replicon mapping technique to show that the replication pattern of the *GNAI3*–*GNAT2* region in cells of lines 474 and 633 is similar to those observed in cells of line 42, which contains only unrearranged and untranslocated copies of the region. Since amplified copies in cells from line 474 are organized as inverted repeats of a truncated region, these results imply that all features essential for replication initiation in the *GNAI3*–*GNAT2* region are located downstream of the

rearrangement site in these cells (Fig. 1). Similarly, the inverted arrangement of amplified units in cells from line 633 clearly demonstrates that features essential for replication initiation are located upstream of the rearrangement site in these cells (Fig. 2). Taken together, our data indicate that the features essential for the function of *oriGNAI3* are located within the 60 kb region separating the two rearrangement sites, and thus favor the notion that local sequences or structures are important for replication initiation at this locus. In agreement with this hypothesis, the pattern of hypersensitivity to DNase I in the 60 kb between the two rearrangement sites reveals a distinctive accessibility of a 12 kb region containing *oriGNAI3* (Fig. 3), which suggests that the function of this origin is mostly specified by this narrow region. These results, which point to local *cis*-acting sequences or structures, may be considered as supporting the hypothesis of mammalian replicators. Replicators, defined as sequences to which initiator proteins bind, are still uncharacterized in mammalian cells. Several structural or sequence motifs have, however, been proposed to be involved in replication initiation, and we show here that such candidate features are clustered in a 7 kb region centered on *oriGNAI3*. Although the biological relevance of each of these features will have to be determined with functional assays, such a clustering of motifs, in the region exhibiting hypersensitivity to DNase I, further support our conclusion that *oriGNAI3* essentially requires local sequence and/or chromatin features. Given that this 7 kb region includes the 3'UTR of the *GNAI3* gene, a potential overlap of the signals important for replication initiation and those specifying transcription termination is suggested, a situation similar to the one observed for several replication origins in *Saccharomyces cerevisiae* (52).

The location of *oriGNAI3* close to the *GNAI3* gene, and the potential overlap of transcription termination and replication initiation features, raised the question of the relationship between transcription and replication at this locus. Indeed, the nature, if any, of such a relationship is poorly understood. Some mammalian DNA replication origins were characterized in long intergenic regions or in transcriptionally silent genes, and several studies suggested that transcription inhibits replication initiation (53,54). On the other hand, other initiation regions have been found to map at, or very close to transcribed sequences (36,44,55–59), which led to the proposal that the two processes are tightly linked in these cases. For example, transcription factors could be directly involved in the replication process, or their binding could alter the chromatin structure in a way that facilitates replication initiation (44,59). However, the recent observation that transcription and replication are temporally distinct processes may have challenged these hypotheses (60). In the present study, we have analyzed the cell line 474, in which amplified *GNAI3* genes have been disrupted by a rearrangement that has deleted promoter sequences (Fig. 1A). This rearrangement has linked amplified units in an inverted orientation, so that no sequence that might compensate for the loss of this promoter has been added (Fig. 1B). As a consequence, amplified truncated *GNAI3* genes are not transcribed, as shown by northern blot analyses (34). PCR analyses indicate here that *oriGNAI3* functions in the absence of transcription of the *GNAI3* gene. These results show that the mapping of a mammalian replication origin at or close to a gene, or the potential overlap of transcription termination and replication initiation features, do not imply that replication initiation relies on active transcription of the gene.

ACKNOWLEDGEMENTS

We would like to thank Professors D. Kowalski and J. N. Anderson for the use of the Thermodyn and GentBen programs, Dr E. Heard for critical reading of the manuscript, and Professor G. Buttin for constant support. This work was supported in part by the University Pierre et Marie Curie, the Ligue Nationale Française contre le Cancer (Comité de Paris), the Association pour la Recherche sur le Cancer, and the Fondation de France.

REFERENCES

- 1 Kobayashi,T., Rein,T. and DePamphilis,M.L. (1998) *Mol. Cell. Biol.*, **18**, 3266–3277.
- 2 Wang,S., Dijkwel,P.A. and Hamlin,J.L. (1998) *Mol. Cell. Biol.*, **18**, 39–50.
- 3 Hu,B., Burkhart,R., Schulte,D., Musahl,C. and Knippers,R. (1993) *Nucleic Acids Res.*, **21**, 5289–5293.
- 4 Gavin,K.A., Hidaka,M. and Stillman,B. (1995) *Science*, **270**, 1667–1671.
- 5 Williams,R.S., Shoet,R.V. and Stillman,B. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 142–147.
- 6 Burhans,W.C., Vassilev,L., Caddle,M.S., Heintz,N.H. and DePamphilis,M.L. (1990) *Cell*, **62**, 955–965.
- 7 Heinzel,S.S., Krysan,P.J. and Calos,M.P. (1991) *Mol. Cell. Biol.*, **11**, 2263–2272.
- 8 Handeli,S., Klar,A., Meuth,M. and Cedar,H. (1989) *Cell*, **57**, 909–920.
- 9 Gilbert,D.M., Miyazawa,H. and DePamphilis,M.L. (1995) *Mol. Cell. Biol.*, **15**, 2942–2954.
- 10 Gilbert,D.M. (1998) *Curr. Opin. Genet. Dev.*, **8**, 194–199.
- 11 Wu,J.R., Keezer,S.M. and Gilbert,D.M. (1998) *EMBO J.*, **17**, 1810–1818.
- 12 Aladjem,M.I., Groudine,M., Brody,L.L., Dieken,E.S., Fournier,R.E., Wahl,G.M. and Epner,E.M. (1995) *Science*, **270**, 815–819.
- 13 Kitsberg,D., Selig,S., Keshet,I. and Cedar,H. (1993) *Nature*, **366**, 588–590.
- 14 Aladjem,M.I., Rodewald,L.W., Kolman,J.L. and Wahl,G.M. (1998) *Science*, **281**, 1005–1009.
- 15 Hamlin,J.L. and Dijkwel,P.A. (1995) *Curr. Opin. Genet. Dev.*, **5**, 153–161.
- 16 Toledo,F., Baron,B., Fernandez,M.A., Lachagès,A.M., Mayau,V., Buttin,G. and Debatisse,M. (1998) *Nucleic Acids Res.*, **26**, 2313–2321.
- 17 Baron,B., Fernandez,M.A., Toledo,F., LeRoscouet,D., Mayau,V., Martin,N., Buttin,G. and Debatisse,M. (1994) *Genomics*, **24**, 288–294.
- 18 Baron,B., Fernandez,M.A., Carignon,S., Toledo,F., Buttin,G. and Debatisse,M. (1996) *Mamm. Genome*, **7**, 922–923.
- 19 Debatisse,M., Toledo,F., Robert de Saint Vincent,B. and Buttin,G. (1992) In Kellems,R.E. (ed.), *Gene Amplification in Mammalian Cells—A Comprehensive Guide*. Marcel Dekker Inc., New York, NY, pp. 173–183.
- 20 Baron,B., Fernandez,M.A., Carignon,S., Toledo,F., Buttin,G. and Debatisse,M. (1996) *Mamm. Genome*, **7**, 429–432.
- 21 Fernandez,M.A., Baron,B., Prigent,M., Toledo,F., Buttin,G. and Debatisse,M. (1997) *J. Cell Biochem.*, **67**, 541–551.
- 22 Toledo,F., LeRoscouet,D., Buttin,G. and Debatisse,M. (1992) *EMBO J.*, **11**, 2665–2673.
- 23 Toledo,F., Buttin,G. and Debatisse,M. (1993) *Curr. Biol.*, **3**, 255–264.
- 24 Toledo,F., Smith,K.A., Buttin,G. and Debatisse,M. (1992) *Mutat. Res.*, **276**, 261–273.
- 25 Dijkwel,P.A., Vaughn,J.P. and Hamlin,J.L. (1991) *Mol. Cell. Biol.*, **11**, 3850–3859.
- 26 Diviacco,S., Norio,P., Zentilin,L., Menzo,S., Clementi,M., Biamonti,G., Riva,S., Falaschi,A. and Giacca,M. (1992) *Gene*, **122**, 313–320.
- 27 Tao,L., Nielsen,T., Friedlander,P., Zannis-Hadjopoulos,M. and Price,G. (1997) *J. Mol. Biol.*, **273**, 509–518.
- 28 RecillasTarga,F., Razin,S.V., deMouraGallo,C.V. and Scherrer,K. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 4422–4426.
- 29 Hyrien,O., Debatisse,M., Buttin,G. and Robert de Saint Vincent,B. (1987) *EMBO J.*, **6**, 2401–2408.
- 30 Dobbs,D.L., Shaiu,W.L. and Benbow,R.M. (1994) *Nucleic Acids Res.*, **22**, 2479–2489.
- 31 Fitzgerald,D.J., Dryden,G.L., Bronson,E.C., Williams,J.S. and Anderson,J.N. (1994) *J. Biol. Chem.*, **269**, 21303–21314.
- 32 Mastrangelo,I.A., Held,P.G., Dailey,L., Wall,J.S., Hough,P.V.C., Heintz,N. and Heintz,N.H. (1993) *J. Mol. Biol.*, **232**, 766–778.
- 33 Rein,T., Zorbas,H. and DePamphilis,M.L. (1997) *Mol. Cell. Biol.*, **17**, 416–426.
- 34 Debatisse,M., Hyrien,O., Petit-Koskas,E., Robert de Saint Vincent,B. and Buttin,G. (1986) *Mol. Cell. Biol.*, **6**, 1776–1781.
- 35 Hyrien,O., Debatisse,M., Buttin,G. and Robert de Saint Vincent,B. (1988) *EMBO J.*, **7**, 407–417.
- 36 Giacca,M., Zentilin,L., Norio,P., Diviacco,S., Dimitrova,D., Contreas,G., Biamonti,G., Perini,G., Weighardt,F., Riva,S. and Falaschi,A. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 7119–7123.
- 37 Weintraub,H. and Groudine,M. (1976) *Science*, **193**, 848–856.
- 38 Kioussis,D., Vanin,E., deLange,T., Flavell,R.A. and Grosveld,F.G. (1983) *Nature*, **306**, 662–666.
- 39 Stamatoyannopoulos,J.A., Goodwin,A., Joyce,T. and Lowrey,C.H. (1995) *EMBO J.*, **14**, 106–116.
- 40 Rao,H. and Stillman,B. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 2224–2228.
- 41 Dimitrova,D.S., Giacca,M., Demarchi,F., Biamonti,G., Riva,S. and Falaschi,A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 1498–1503.
- 42 Bird,A. (1986) *Nature*, **321**, 209–213.
- 43 Larsen,F., Gundersen,G., Lopez,R. and Prydz,H. (1992) *Genomics*, **13**, 1095–1107.
- 44 Delgado,S., Gomez,M., Bird,A. and Antequera,F. (1998) *EMBO J.*, **17**, 2426–2435.
- 45 Natale,D.A., Umek,R.M. and Kowalski,D. (1993) *Nucleic Acids Res.*, **21**, 555–560.
- 46 Lin,S. and Kowalski,D. (1997) *Mol. Cell. Biol.*, **17**, 5473–5484.
- 47 Gasser,S.M. (1991) *Curr. Opin. Cell Biol.*, **3**, 407–413.
- 48 Maric,C. and Hyrien,O. (1998) *Chromosoma*, **107**, 155–165.
- 49 Pemov,A., Bavykin,S. and Hamlin,J.L. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14757–14762.
- 50 Williams,J.S., Eckdahl,T.T. and Anderson,J.N. (1988) *Mol. Cell. Biol.*, **8**, 2763–2769.
- 51 Marahrens,Y. and Stillman,B. (1992) *Science*, **255**, 817–823.
- 52 Chen,S., Reger,R., Miller,C. and Hyman,L.E. (1996) *Nucleic Acids Res.*, **24**, 2885–2893.
- 53 Brewer,B.J. (1994) *Curr. Opin. Genet. Dev.*, **4**, 196–202.
- 54 Haase,S.B., Heinzel,S.S. and Calos,M.P. (1994) *Mol. Cell. Biol.*, **14**, 2516–2524.
- 55 Ariizumi,K., Wang,Z. and Tucker,P.W. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3695–3699.
- 56 Zhao,Y., Tsutsumi,R., Yamaki,M., Nagatsuka,Y., Ejiri,S. and Tsutsumi,K. (1994) *Nucleic Acids Res.*, **22**, 5385–5390.
- 57 Gale,J.M., Tobey,R.A. and D’Anna,J.A. (1992) *J. Mol. Biol.*, **224**, 343–358.
- 58 Tasheva,E.S. and Roufa,D.J. (1994) *Mol. Cell. Biol.*, **14**, 5628–5644.
- 59 Kelly,R.E., DeRose,M.L., Draper,B.W. and Wahl,G.M. (1995) *Mol. Cell. Biol.*, **15**, 4136–4148.
- 60 Wei,X., Samarabandu,J., Devdhar,R.S., Siegel,A.J., Acharya,R. and Berezney,R. (1998) *Science*, **281**, 1502–1505.