 Nucleic Acids Research, 1999, Vol. 27, No. 15 **3009–3017**

An origin of bidirectional DNA replication is located within a CpG island at the 3' end of the chicken lysozyme gene

Loc Phi-van* and Wolf H. Strätling[1]

Institut für Tierzucht und Tierverhalten, Bundesforschungsanstalt für Landwirtschaft, Dörnbergstrasse 25–27, 29223 Celle, Germany and [1]Institut für Physiologische Chemie, Universität Hamburg, Martinistrasse 52, 20246 Hamburg, Germany

Received May 28, 1999; Accepted June 16, 1999

ABSTRACT

We previously identified a broad initiation zone of DNA replication at the chicken lysozyme gene locus. However, the existence of a highly preferred origin of bidirectional replication (OBR), often found in initiation zones, remained elusive. In order to re-examine this issue we used a competitive PCR assay to determine the abundance of closely spaced genomic segments in a 1 kb size fraction of nascent DNA. A sharp peak of nascent strand abundance occurred at the 3' end of the gene, where initiation events were 17 times more frequent than upstream of the gene. This primary initiation site, active in lysozyme expressing myelomonocytic HD11 cells and non-expressing hepatic DU249 cells, was found to reside within an unusually located CpG island. While most CpG islands are found at the 5' end of genes, the lysozyme gene island extends from the 3' end of the second intron and includes ~1.2 kb of 3' flanking DNA. As diagnosed by methylation-sensitive restriction enzymes, the island is largely non-methylated in HD11 cells, DU249 cells and inactive chicken erythrocytes. Furthermore, a DNase I hypersensitive site (HS) that is composed of two subsites separated by ~100 bp, was localised very close to the segment with the highest initiation activity. Our results suggest that the non-methylated CpG island and the HS provide an accessible chromatin structure for the lysozyme gene origin of replication.

INTRODUCTION

Mapping initiation sites for DNA replication in the chromosomes of flies, frogs and mammals has revealed at least 22 replication origins (reviewed in 1 and references cited below). While these studies agree that DNA synthesis initiates at specific sites but not randomly throughout cellular chromosomes, an important feature of replication origins appears paradoxical. Several studies using labelling of early replication intermediates with nucleotide precursors conclude that initiation occurs from narrow zones encompassing as little as 0.5 kb (2–4). In contrast, when replication intermediates were fractionated by two-dimensional gel electrophoresis and then identified by hybridisation with sequence-specific probes, initiation appears to occur at many sites distributed throughout much larger regions. In several cases this paradox could be clarified by use of techniques that allow quantitative determination of the relative frequency of initiation at densely spaced sites (4,5). The chicken lysozyme gene locus is one of the best studied chromatin domains in vertebrates. The gene is flanked by well characterised regulatory sites (enhancers, silencer, HRE) (6) and nuclear matrix attachment regions (MARs) (7). To fill a gap in our knowledge, we started to study replication of the lysozyme locus, and as a first step, previously identified a broad initiation zone of DNA replication (8). Yet unfortunately, our previous studies failed to reveal a preferred origin of bidirectional replication (OBR) in the lysozyme initiation zone. We therefore decided to re-examine this issue and used the nascent-strand abundance assay to quantitatively determine by PCR the relative abundance of more densely spaced genomic sites within the 1.0 kb size fraction of nascent DNA from chicken myelomonocytic HD11 cells and hepatic DU249 cells. Furthermore, we concentrated our efforts on the 3' half of the gene and its 3' flanking DNA, as our previous studies seemed to indicate that the frequency of initiation is greatest within the 3' located sub-zone (8).

Presently very little is known about the distribution of origins of DNA replication in relationship to other genomic landmarks. Origins were found to occur near promoters (3,9), within transcriptional units (10), at the 3' end of genes (3) and in intergenic regions (2,4,11). A close correlation of DNA replication with transcription is demonstrated by the fact that almost all constitutively expressed (housekeeping) genes replicate within the first half of S phase in many cell types, while many tissue specific genes when active replicate in early S phase, but when inactive in late S phase (12,13). A very prominent landmark within the vertebrate genome is 1–2 kb regions of densely spaced CpG dinucleotides, named CpG islands. They are associated with the 5' end of ~50% of all mammalian genes, where they cover the promoter and one or more of the 5' exons (14,15). While the majority of the methylated CpGs are found in retrotransposons and centromeric satellite DNA, CpG islands are free of methylation in all tissues, including those in which the associated genes are silent, with the exception of

*To whom correspondence should be addressed. Tel: +49 5141 384646; Fax: +49 5141 381849; Email: loc.phi-van@fal.de

those on the inactive X chromosome (16) and those associated with the inactive allele of imprinted genes (17,18). The chromatin encompassing CpG islands is free of histone H1, highly acetylated at histones H3 and H4 N-termini, and preferentially accessible to restriction enzymes (19). These features provide an 'open' chromatin configuration. In a recent nascent-strand abundance analysis of three hamster genes and one human gene, that all contain CpG islands at their 5' ends, CpG island DNA but not flanking DNA was found to be present in a short nascent strand size fraction (20). Furthermore, CpG island-like fragments were enriched in the short nascent DNA fraction from human erythroleukaemic cells, suggesting that initiation occurs in CpG islands at a significant frequency. In order to evaluate the generality of this connection, it would be interesting to determine the origin of replication in genomic loci that contain CpG islands at non-5' end regions.

Besides by DNA sequences, origins of replication are determined by chromatin structure. Initiation of DNA replication is strongly inhibited by packaging of DNA into chromatin. This effect can be abolished by interaction of specific proteins, e.g. initiation proteins and some transcription factors, with replication origins (21,22). The origin of simian virus 40 mini-chromosome and origins of *Saccharomyces cerevisiae* have been shown to be maintained in an open chromatin structure highly accessible to DNA endonucleases (23–25). Furthermore, the chromatin structure at origins in yeasts and origins in human cells is modulated during the cell cycle (23,26,27). Additionally, it has been shown that attachment to the nuclear matrix mediates cell cycle-specific alterations in chromatin structure at the hamster DHFR ori-β and ori-γ (28).

Our present results localise a primary initiation site (OBR) to a 650 bp region at the 3' end of the lysozyme gene. This OBR resides within an ~1.8 kb CpG island, which extends from the 3' end of the second intron and includes ~1.2 kb of 3' flanking DNA. Furthermore, the OBR is coincident with two closely spaced DNase I hypersensitive sites. These results appear to suggest that the CpG island and the hypersensitive sites establish an 'open' chromatin configuration, which is important for the function of the lysozyme gene 3' end as an OBR.

MATERIALS AND METHODS

Cell culture and cell synchronization

HD11 cells (29) and DU249 cells (30) were maintained in Iscove's modified Dulbecco's medium (IMDM), supplemented with 8% fetal calf serum, 2% chicken serum, 100 U/ml penicillin and 100 µg/ml streptomycin (complete IMDM) at 37°C and 5% CO_2. For synchronisation, HD11 cells (2×10^7) were cultured in 150 mm diameter dishes to 80–90% confluence, then incubated for 45 h in IMDM without isoleucine, and finally transferred to complete IMDM containing 20 µg/ml aphidicolin (Sigma, Deisenhofen, Germany) for at least 12 h (31). To allow these cells, arrested at the G_1/S boundary, to enter S phase, they were washed twice in complete IMDM to remove aphidicolin, and immediately released into S phase by incubation at 37°C for 1 h. For accumulation in G_0 phase, HD11 cells (2×10^7) were grown in complete IMDM overnight and then deprived of serum for at least 72 h. To re-enter G_1 phase, these cells were washed once with IMDM and incubated in complete IMDM for 3 h. Chicken erythrocytes were collected as follows. Heparinised (10 ml) blood was diluted with 10 ml IMDM and centrifuged through 10 ml of Ficoll-Paque (Pharmacia, Freiburg, Germany) at 400 *g* and 4°C for 30 min, followed by two washes with phosphate-buffered saline (PBS).

Isolation and size fractionation of nascent DNA

In order to isolate nascent strand DNA for competitive PCR analysis, 4×10^8 HD11 cells grown in complete IMDM overnight were pulse-labelled with 5-bromodeoxyuridine (BrdUrd, 20 µM) for 15 min at 37°C. High molecular weight DNA was isolated by digestion with proteinase K, extraction with phenol-chloroform and spooled in 70% ethanol. Then, spooled DNA was dissolved in 500 µl of 0.2 N NaOH, layered on two 5–30% (w/v) linear alkaline sucrose gradients containing 0.2 N NaOH and 2 mM EDTA and size-fractionated by centrifugation in a Beckman SW40 rotor at 35 000 r.p.m. for 18 h at 15°C. From each gradient, 12 fractions of 1 ml each were collected, neutralised with 0.2 N HCl and 0.1 M Tris–HCl (pH 8.0) and precipitated with ethanol. DNA sizes in each fraction were monitored by alkaline agarose gel electrophoresis and Southern hybridization to ^{32}P-labelled, nick-translated genomic DNA using *Hin*dIII-digested λDNA and a 400 bp DNA ladder as size markers. From each 1 kb size fraction (range, 0.8–1.2 kb) nascent DNA was purified by two cycles of immunoprecipitation exactly as described by Vassilev and Johnson (9) using 30 µl each of an anti-BrdUrd monoclonal antibody (25 µg/ml; Becton-Dickinson, Heidelberg, Germany), and dissolved in 200 µl TE [10 mM Tris–HCl (pH 8.0), 1 mM EDTA].

Oligonucleotide primers and competitive PCR

The purified nascent DNA was analysed by competitive PCR at 10 segments of the lysozyme gene locus (Fig. 1). For each segment, a primer set, consisting of two external primers (primers 1 and 2) and two internal primers (primers 3 and 4), was chemically synthesised by MWG-Biotech (Ebersberg, Germany). The nucleotide sequences of the primers are given in Table 1. Each internal primer carries at its 5' end a 20 nt sequence unrelated to chicken genomic DNA, either 5'-acct-gcagggatccgtcgac-3' (tail 1) or 5'-gtcgacggatccctgcaggt-3' (tail 2). Specific competitors, that have the same sequence as the genomic segments but in addition the 20 nt sequence in the middle, were constructed by PCR as described previously (8,32). Concentrations of the competitors were determined by co-amplification with a known amount of chicken genomic DNA.

Competitive PCRs were performed in 50 µl reaction mixtures containing 1× GeneAmp buffer, 1.25 U AmpliTaq Gold polymerase (Perkin-Elmer, Hamburg, Germany), four deoxyribonucleotides (200 µM each), 1 µM primers, 5 µl of 1 kb nascent DNA and increasing amounts of specific competitor DNA. Forty-three cycles of amplification, each consisting of 1 min at 94°C and 1 min at 60°C were performed in a Perkin-Elmer thermal cycler. Amplified DNA fragments were fractionated on 5% polyacrylamide gels at 150 V for 2 h, stained with ethidium bromide and quantified by using an ethidium bromide gel documentation system from Bio-Rad (Munich, Germany).

DNA methylation analysis

To isolate genomic DNA for DNA methylation analysis, HD11 and DU249 cells, synchronised HD11 cells and erythrocytes

Table 1. List of primer sequences

Name	Sequence (5' to 3')
B/1	AGAGCGATGCTCAGTAAGGC
B/2	ATGCAGCTTGCTTCCTATGC
B/3	GCCAAAGAGTCTGCTGAATG
B/4	CTGCTGGAATCAGGAAACTG
C/1	TCTTCCATGTTGGTGACAGC
C/2	ATCAATCCATGCCAGTAGCC
C/3	GAGGTCAAGTTACGAACTCA
C/4	AGCTGGGGTCAATAAGTAAC
D/1	CAGTCGTGGAGTTGTATGCG
D/2	ATACAAGCAGCAATCTGGCC
D/3	ATTTCAAGGAGAATGGATCG
D/4	AAACCAGTACATACCCATAG
I/1	ACGCATCAGTGGACGGTTTAC
I/2	CAACACCCACCTCGGAGAG
I/3	CTACGAGAGAAGCTCGCTGT
I/4	GACATGTCCGCCTACGTGAA
K/1	GCAACACTTGGCAAACCTCAC
K/2	ACTACACGGCCTTCAGCACAG
K/3	ATCACTGGGGTCAGGACAGT
K/4	GACCGTGTCCCACCTAAAGC
M/1	CCACTAGTGAAGGGGAGGAGA
M/2	AGTGCAGCTGCCAGAATACC
M/3	CTTAATCTAACCAAGGGGGA
M/4	GTGAAACAACACTCATGGTC
N/1	TGAGAGGGGGTTGGGTGTAT
N/2	CGCTCTACGCATTCTGAAACA
N/3	CGTGGCATAGTGCCAGCAGT
N/4	GCCTGCTTGTGACTCTGAGA
O/1	CATAACAGCGAGCGTGAACTG
O/2	ACGTCGGTGCCCTTGC
O/3	CGCCACCTACCACGCGTT
O/4	GGGGGTTCCCAGGAGA
P/1	ACGACACTGGCAACATGAGG
P/2	ATTCCAACATCACGCAGACC
P/3	AGTTATCAAGTCCGTGACGC
P/4	ATCGGGGATACAGCCTGGGA
R/1	GCTGCACAGCCGGCCGCTTTG
R/2	CCCGCCCGGCCCCCTTCC
R/3	CGGACCGACGGGAAGGCGTT
R/4	CGTCTCTTTCCGCCGCCAGG

(~4×10^7 cells) were washed twice with PBS, lysed in 3 ml of a solution containing 0.3 M NaCl, 50 mM Tris–HCl (pH 8.0), 25 mM EDTA, 0.2% (w/v) SDS and 0.2 mg/ml proteinase K. Following incubating at 37°C for 5 h, the lysate was extracted three times with Roti Phenol (Roth, Karlsruhe, Germany). DNA was spooled from 70% ethanol, washed once with 70% ethanol and dissolved in 2 ml TE. The sample was then digested with 50 µg/ml DNase-free RNase A in the presence of 10 mM EDTA and 150 mM NaCl for 3 h at 37°C, followed by a digestion with 0.2 mg/ml proteinase K for 2 h at 37°C. After extraction, DNA was then precipitated with ethanol, washed and dissolved in 200 µl TE.

For Southern analysis, 10 µg of purified DNA was digested with the indicated restriction enzymes, and the resulting DNA fragments were electrophoretically resolved on 1, 1.4 or 2% agarose gels containing 1× TBE [89 mM Tris (pH 8.0), 89 mM boric acid and 2 mM EDTA], and transferred onto nylon membranes by the method of Southern (33). Filters were hybridised to ^{32}P-labelled, nick-translated DNA probes as described previously (34) and autoradiographed at –80°C using intensifying screens.

Mapping of DNase I hypersensitive (HS) sites

Approximately 1×10^8 cells washed twice in PBS were scraped in 20 ml of buffer A [15 mM NaCl, 60 mM KCl, 15 mM Tris–HCl (pH 7.4), 2 mM EDTA, 0.5 mM EGTA, 0.15 mM spermidine, 0.5 mM spermine, 0.5 M sucrose, 0.5% Triton X-100 and 1 mM PMSF] (35) with a rubber policeman, and homogenized in a Dounce homogenizer (type A pestle). Following centrifugation at 2000 *g* for 10 min at 4°C, nuclei were washed twice in 10 ml of the same buffer, twice in 10 ml of buffer B [15 mM NaCl, 60 mM KCl, 15 mM Tris–HCl (pH 7.4), 2 mM EDTA, 0.5 mM EGTA, 0.15 mM spermidine, 0.5 mM spermine, 0.35 M sucrose and 1 mM PMSF], and then suspended in buffer C [15 mM NaCl, 60 mM KCl, 15 mM Tris–HCl (pH 7.4), 0.2 mM EDTA, 0.2 mM EGTA, 0.15 mM spermidine, 0.5 mM spermine and 1 mM PMSF] at a nucleic acid concentration of 1 mg/ml.

To analyse hypersensitive sites, nuclei (200 µg of nucleic acids) were treated without DNase I or with 50–150 U/ml of DNase I in 600 µl buffer C in the presence of 5 mM $MgCl_2$ at 4°C for 10 min. Reactions were stopped by adding 12.5 µl of 0.5 M EDTA. After centrifugation for 10 s at 12 000 r.p.m., DNA from digested nuclei was purified as described previously (34). Following digestion with *Xba*I and *Hin*dIII, DNase I HS sites were mapped by the technique of indirect end labelling (36) using DNA fragments B4B5 and CeH10 as probes. Southern hybridization was performed as described previously (34).

DNA sequencing

DNA fragments to be sequenced were cloned into pBluescript II SK+ (Stratagene Europe, Amstersdam, The Netherlands). DNA sequencing was performed by using a T7 sequenase kit from USB (Amersham Buchler, Braunschweig, Germany) with universal sequencing primers according to the instructions of the manufacturer.

RESULTS

Identification of an OBR in the chicken lysozyme gene locus

We have previously mapped an initiation zone of DNA replication at the chicken lysozyme gene locus and presented evidence that it contains a 3' located sub-zone of preferred initiation (8). In the present study, we wished to determine

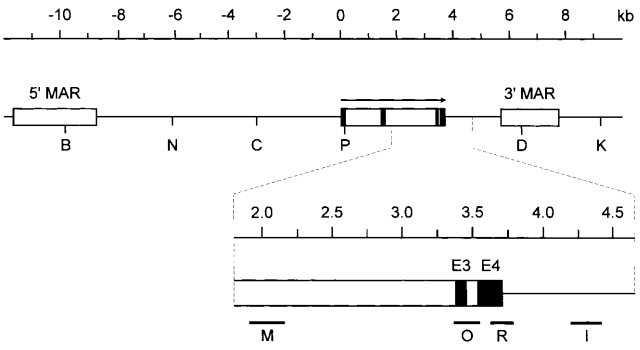

Figure 1. Schematic representation of the chicken lysozyme gene locus and the segments analysed. The lysozyme gene, its exons (E), and the direction of transcription are represented by the central box with vertical bars and a horizontal arrow. The region around the 3' end of the gene is enlarged in the lower part of the figure. Flanking MARs are shown by open boxes. The relative location of the segments analysed is indicated below the map. Segments B, C, D, K, M and N in this study are the same sites as lys B, lys C, lys D, lys I, lys H and lys G (8), respectively. Primer sequences are provided in Table 1. The upper line gives the distances (kb) from the 5' end of the lysozyme gene.

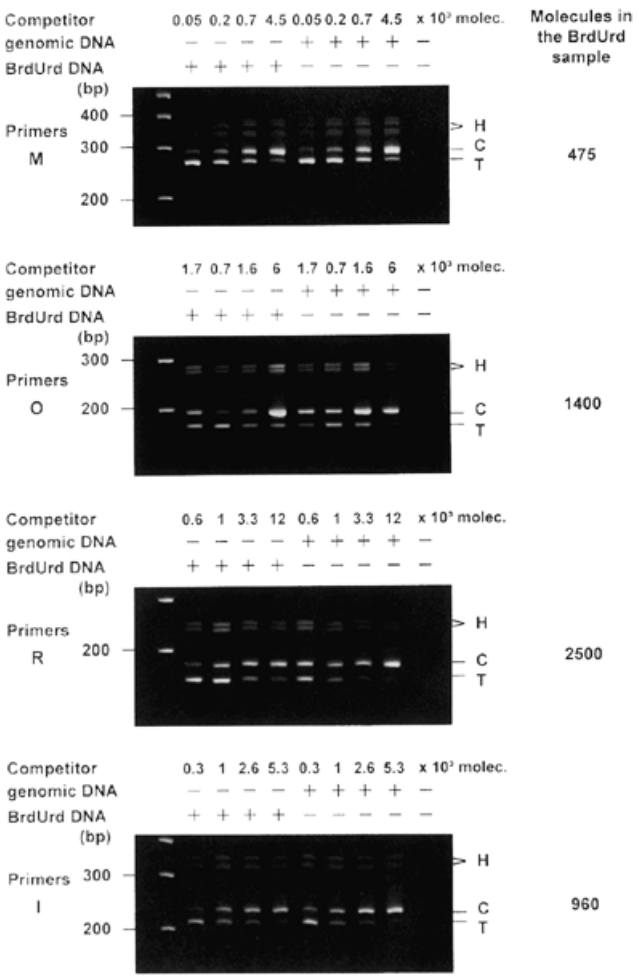

Figure 2. Competitive PCR on the 1 kb nascent strand size fraction. A fixed amount of the 1 kb nascent strand size fraction or 2 ng genomic DNA, corresponding to approximately 600 genomic equivalents, were co-amplified in the presence of increasing amounts of specific competitors using primer sets M, O, R and I. PCR products were electrophoretically resolved on 5% polyacrylamide gels and stained with ethidium bromide. Marker fragments were a 100 bp ladder. T and C bands contain target and competitor PCR products, respectively. Two heteroduplex DNA bands consisting of either T.C or C.T are indicated by H. The number of molecules of M, O, R and I calculated for the 5 µl sample of BrdUrd-labelled DNA is shown on the right.

using more closely spaced genomic markers, whether this subzone harbours an OBR. Employing the competitive PCR-based nascent strand assay, we measured the abundance of ten genomic segments (B, N, C, P, M, O, R, I, D and K) in a 1 kb size fraction of BrdUrd-labelled, nascent strand DNA. The location of these segments is shown in Figure 1. By spanning ~20 kb, they cover the 21 kb lysozyme gene domain (6,7,37). Four segments, M, O, R and I, located in the preferred initiation zone are spaced by 1140, 55 and 430 bp, respectively. The distances between these probes are small enough to allow detection of an OBR in this region. The competitive PCR assay was performed by co-amplification of a fixed amount of the BrdUrd-labelled nascent strand DNA together with varying known amounts of competitor DNA. The results of a representative experiment with segments M, O, R and I are shown in Figure 2. The accuracy of this determination was validated by parallel competitive PCR assays of a fixed amount of unlabelled, unfractionated genomic DNA with varying amounts of competitor DNA. We found that the 2 ng genomic DNA sample used contained an average of 600 ± 106 copies (standard error of the mean) for each genomic marker (Fig. 2, right half of each panel). Consistent with the observations of Kobayashi *et al.* (4), not only competitor (C) and target (T) DNA bands appeared as PCR products, but also heteroduplexes (H) between target and competitor DNA amplicons, T.C and C.T, that formed during the final cycle of denaturation and renaturation. The amount of this heteroduplex DNA formed was taken into account using the formula given by Kobayashi *et al.* (4). Then, from the ratio of competitor to target DNA and the number of competitor DNA molecules added, we calculated the number of copies of each genomic segment in the 1 kb nascent DNA size fraction. In total, we performed three independent measurements for each genomic segment, and average copy numbers were normalised to the average copy number of segment B (Fig. 3A). The average abundance of segment R, located at the 3' end of the lysozyme gene, was (17.0 ± 1.8)-times higher than that of segment B. The average enrichment of segments O and I, which flank segment R, was (10.5 ± 1.0)-fold and (6.5 ± 0.6)-fold, respectively. Segments M and D, which are 1445 bp upstream and 2539 bp downstream from segment R, show a (3.2 ± 0.5)- and (1.9 ± 0.2)-fold average enrichment, respectively. Thus the normalised abundance of nascent DNA exhibits a symmetrical peak centred at segment R, indicating bidirectional replication from a site within the ~650 bp region between segments O and I. The distance from the 5' end of segment O to the 3' end of segment I measures 1041 bp. Nascent strands were found to be ~1.6- and ~2.6-fold more abundant in segment R than in the immediate upstream (O) and downstream (I) segment, respectively, suggesting that the centre of the OBR resides in the left half of the 1041 bp region. Altogether,

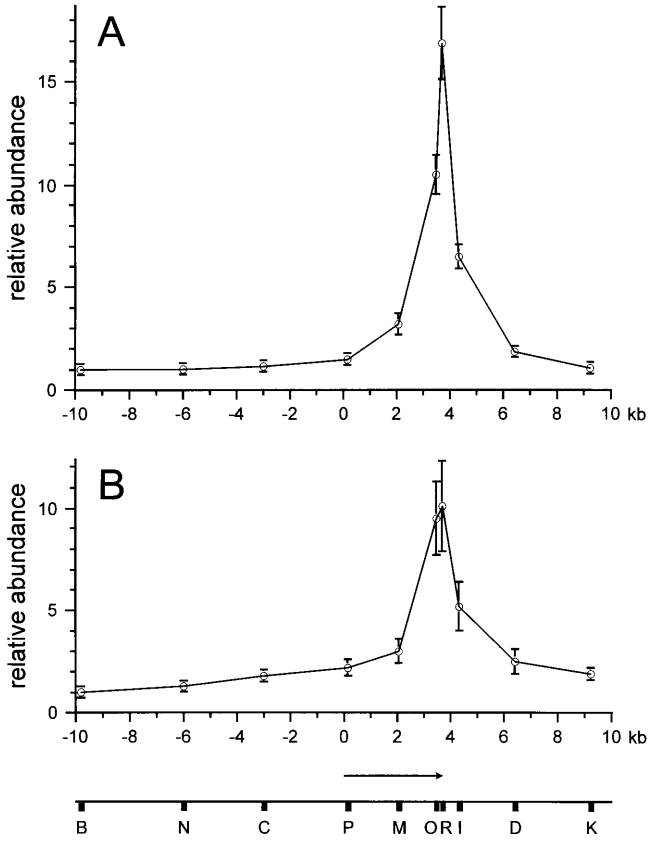

Figure 3. Mapping of an OBR with early replicating nascent DNA. The relative abundance of 10 segments (see Fig. 1 and map at the bottom) was determined on three different samples of nascent, BrdUrd-labelled DNA with an average strand length of ~1.0 kb prepared from myelomonocytic HD11 cells (**A**) and from hepatic DU249 cells (**B**) by competitive PCR. The abundance of each segment was normalised to the average of segment B. The mean values with standard errors are shown.

these data demonstrate an OBR close to the 3' end of the chicken lysozyme gene.

To determine whether this OBR is also active in lysozyme non-expressing hepatic DU249 cells, we performed a competitive PCR analysis of 1 kb nascent strand DNA from these cells with the same 10 pairs of primers. As shown in Figure 3B, the results were very similar to those obtained with nascent strand DNA from HD11 cells, although segment R was enriched ~10-fold over segment B in nascent strand DNA and the flanking segments O and I showed a 9.5- and 5.2-fold enrichment, respectively. These data are consistent with an active OBR close to segment R in DU249 cells.

The lysozyme OBR is located within a CpG island

Sequencing of cloned genomic DNA between segments M and K revealed that a CpG island (66% GC-rich), ~1.8 kb in length, is located at the 3' end of the lysozyme gene. The density of CpG doublets in the central portion of the island is, on average, ~16/100 bp and on its flanks ~7/100 bp, while this density is as low as 0.5–1.5/100 bp upstream and downstream of the island (Fig. 4). The central portion of the island extends from exon 3 and includes ~700 bp of 3' flanking DNA. These data demonstrate that the 1041 bp region (from segment O to segment I) harbouring the identified OBR co-localises with the central portion of the CpG island.

The lysozyme CpG island is undermethylated in myelomonocytic HD11 cells, in hepatic DU249 cells and in chicken erythrocytes

To assess the methylation status of the CpG island in various cell lines, we utilised restriction enzymes, whose activity is inhibited by CpG methylation within their recognition sites. As probes, purified DNA from lysozyme expressing HD11 cells, non-expressing hepatic DU249 cells and transcriptionally and replicationally inactive chicken erythrocytes were digested with *Hpa*II, *Msp*I, *Ava*I, *Bss*HII or *Nar*I, and additionally with *Sac*I. The recognition sites of these enzymes within the CpG island are plotted in Figure 4. Digested DNA samples were analysed by Southern blotting and hybridization with probe S6S7. Figure 5A shows a Southern analysis of the DNA samples digested by *Hpa*II or *Msp*I. Both enzymes recognise the same sequence, CCGG; however, in contrast to *Hpa*II, that is inhibited when the internal cytosine is methylated, *Msp*I is not sensitive to methylation. Although 19 restriction fragments, from 5 to 343 bp in length, should be generated from fragment S6S7 by digestion with *Msp*I, only three bands, 144, 211 and 343 bp in size, were detectable (Fig. 5A, lanes 3, 5, 7 and 9). The low blotting and hybridization efficiency of the many smaller fragments probably impeded their detection. The cleavage pattern after *Hpa*II digestion was very similar to that by *Msp*I, indicating that most CCGG sites are unmethylated (lanes 2, 4, 6 and 8). Yet we noticed one difference: *Hpa*II digestion produced a 393 bp instead of a 343 bp fragment, indicating that the far right-hand CCGG in fragment S6S7 was methylated in exponentially growing HD11 cells (lanes 2 and 3). Identical results were obtained with DNA from HD11 cells that were deprived of serum growth factors (Fig. 5A, lanes 4 and 5, and B, lanes 1 and 2), arrested in G_1 phase (Fig. 5B, lanes 3 and 4), or synchronised into S phase (Fig. 5B, lanes 5 and 6). Additionally, the far right-hand CCGG site outside of fragment S6S7 was found to be methylated in HD11 cells, but not in DU249 cells and erythrocytes (data not shown). Digestion of DNA from HD11 cells, DU249 cells or erythrocytes with methylation-sensitive *Ava*I generated three fragments, of which the largest and the smallest ones had sizes of 692 and 193 bp, respectively (Fig. 5C, lanes 2, 5, 8 and 11). This indicates that the two central cleavage sites are unmethylated (Fig. 4). However, the middle fragment cleaved from exponentially growing or arrested HD11 cell DNA exhibited a size of 424 bp (Fig. 5C, lanes 2 and 5), while the respective fragment cleaved from DU249 cell or erythrocyte DNA was 373 bp in length (Fig. 5C, lanes 8 and 11). These results and the fact that the far right *Ava*I site (CCCGGG) in fragment S6S7 also contains the far right *Msp*I site, confirm our earlier conclusion that the CpG within this common site is methylated. Digestion with *Bss*HII generated two fragments, and that with *Nar*I generated three fragments, of which the two smaller ones were poorly resolved (Fig. 5C, lanes 3, 4, 6, 7, 9, 10, 12 and 13). This indicates that the respective cleavage sites were unmethylated. Taken together, our results suggest that the central portion of the CpG island containing the lysozyme gene OBR is largely unmethylated in the three cell types investigated. In contrast, two CpG sites at the 3' border

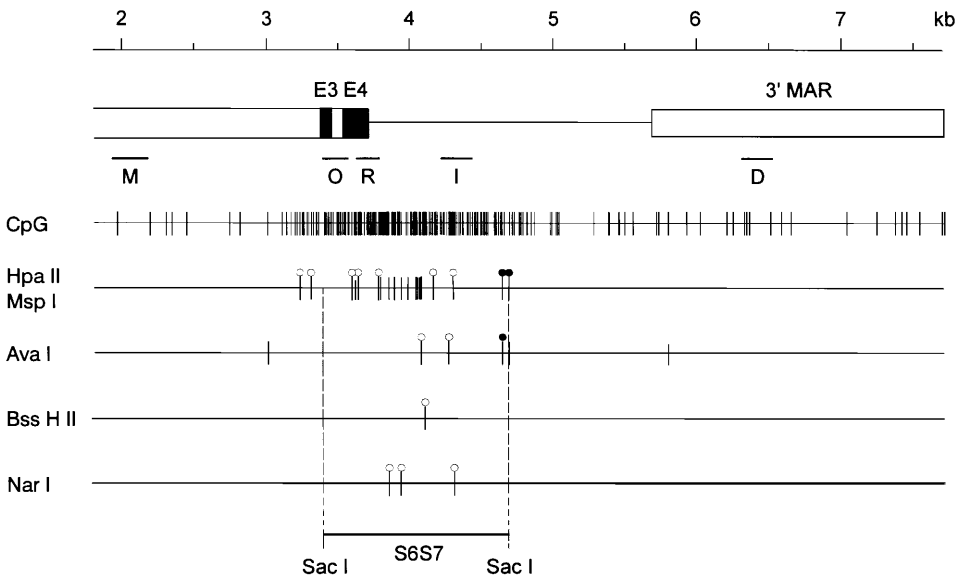

Figure 4. A CpG island is located at the 3' end of the lysozyme gene. Map plot of the 3' half of the lysozyme gene and 3' flanking sequence, indicating introns by open boxes and exons (E3 and E4) by closed boxes, and the 3' MAR by a rectangle. Horizontal bars mark segments M, O, R, I and D. The plot shows the cluster of CpG dinucleotides at the 3' end of the lysozyme gene, and the relative positions of the methylation-sensitive restriction enzyme sites used in the methylation analysis. Unmethylated sites resolved by electrophoresis are indicated by open lollipops. The two unmethylated CCGG sites at the 5' border of the island were analysed by *Hin*dIII digestion and hybridisation to probe S6S7 and a probe encompassing segment M (data not shown). Two adjacent CCGG sites methylated in HD11 cells but not in DU249 cells or erythrocytes are marked by closed lollipops. Analysis of the downstream region of one of these sites used digestion with *Hin*dIII instead of *Sac*I (data not shown). A bar at the bottom and two vertical dotted lines indicate the positions of the S6S7 probe and the flanking *Sac*I sites used.

of the island were found to be methylated in lysozyme expressing HD11 cells.

The lysozyme OBR contains a DNase I HS site

An OBR might be expected to be preferentially accessible to the replication machinery, and thus be packaged in a non-canonical, open chromatin structure. In order to detect such features, we attempted to map sites potentially hypersensitive to DNase I using the technique of indirect end-labelling (36). Nuclei were isolated from exponentially growing HD11 and DU249 cells and incubated with increasing concentrations of DNase I. Following digestion of the purified DNA with *Xba*I and *Hin*dIII, DNase I HS sites were then mapped by indirect end-labelling from sites X4 and H10, respectively (see map in Fig. 6A). As shown in Figure 6B, *Xba*I digestion and hybridisation with probe B4B5 localised a DNase I HS site very close to the 3' end of the lysozyme gene. This site is likely identical with a DNase I HS site previously mapped at position +3.9 kb (named HS6) in various chicken tissues (38). A fine mapping by digestion with *Hin*dIII and hybridization with probe CeH10 then revealed that the HS site detected in Figure 6B consists of two sub-sites spaced by ~100 bp (Fig. 6C and D). Interestingly, these sub-sites are present in both HD11 and DU249 cell lines and map within segment R, that has been earlier shown to display the highest frequency of initiation (Fig. 3). Thus our results demonstrate that a DNase I HS site, which is indicative of a localised accessible chromatin structure, occurs within the lysozyme OBR. It is tempting to postulate that the specialised chromatin organisation of this DNase I HS site plays an important role in the function of the lysozyme OBR.

DISCUSSION

We previously identified an initiation zone of DNA replication within the chicken lysozyme gene locus, but failed to localise a preferred primary initiation site. Re-examining this issue, we used a competitive PCR assay to determine the abundance of selected genomic segments in a 1 kb size fraction of nascent strand DNA. The genomic segments chosen were particularly densely spaced in the 3' half of the gene and its 3' flanking region, since our previous measurements seemed to indicate that initiation events are enhanced in a 3' located sub-zone (8). Using this modification, the present study discovers an OBR at the 3' end of the gene. This OBR localises to an ~650 bp sequence, that contains the last (fourth) exon of the gene and ~400 bp of 3' flanking DNA. Segment R within the OBR was 17-fold enriched in the 1.0 kb size fraction of BrdUrd-labelled nascent DNA. Segments O and I, that are separated from segment R by 55 and 430 bp, respectively, exhibited only a 10.5- and 6.5-fold enrichment, respectively. The sharpness of the peak of nascent strand abundance thus indicates that the primary initiation events localise to a strikingly narrow region very close to the 3' end of the lysozyme gene. We furthermore considered the possibility that an additional high-frequency initiation site would exist at the gene locus. In case of its existence, the flanking segments should exhibit an elevated abundance in the 1 kb nascent strand DNA fraction, since the distance between the segments used ranges from 3 to 4 kb. As shown in Figure 3, solely segments D and M, but not segments N, C, P and K, were found to be enriched. Hence, it is unlikely that other additional high-frequency initiation sites are present at the gene locus.

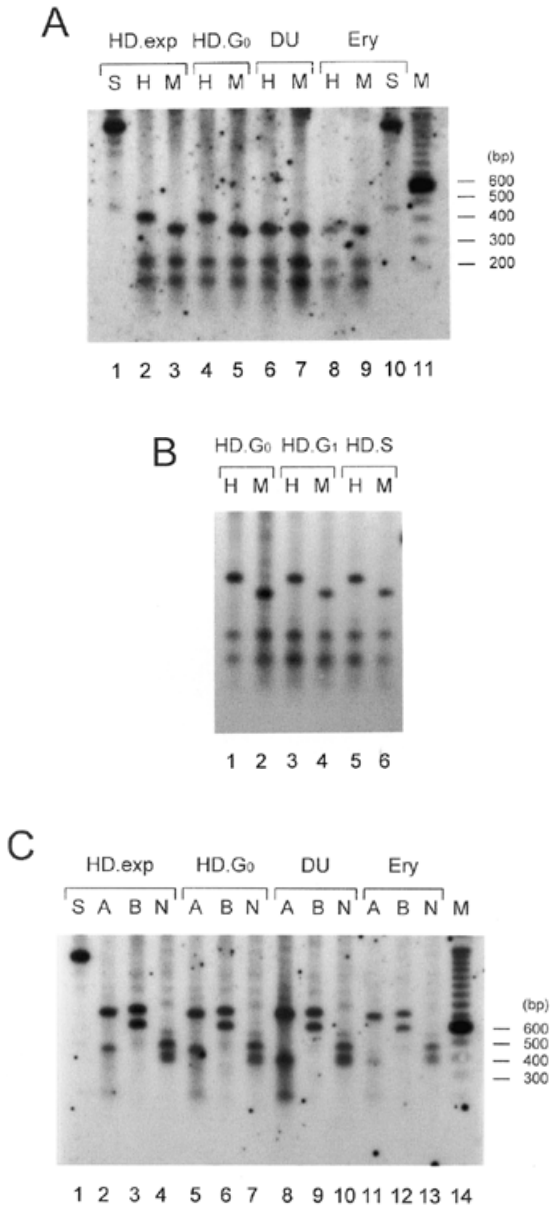

Figure 5. DNA methylation analysis of the lysozyme CpG island. Genomic DNA was isolated from exponentially growing HD11 cells (HD.exp), HD cells deprived of serum (HD.G_0), HD11 cells arrested in G_1 (HD.G_1) or S phase (HD.S), DU249 cells (DU) or chicken erythrocytes (Ery), and digested with *Sac*I (S) or *Sac*I in addition to *Hpa*II (H), *Msp*I (M), *Ava*I (A), *Bss*HII (B) or *Nar*I (N). Digested DNA samples were resolved on 2% (**A** and **B**) or 1.4% agarose gels (**C**), Southern-blotted and hybridised to probe S6S7 (Fig. 4). Lanes M (A and C) contain a 100 bp DNA ladder and marker sizes are noted on the right-hand side.

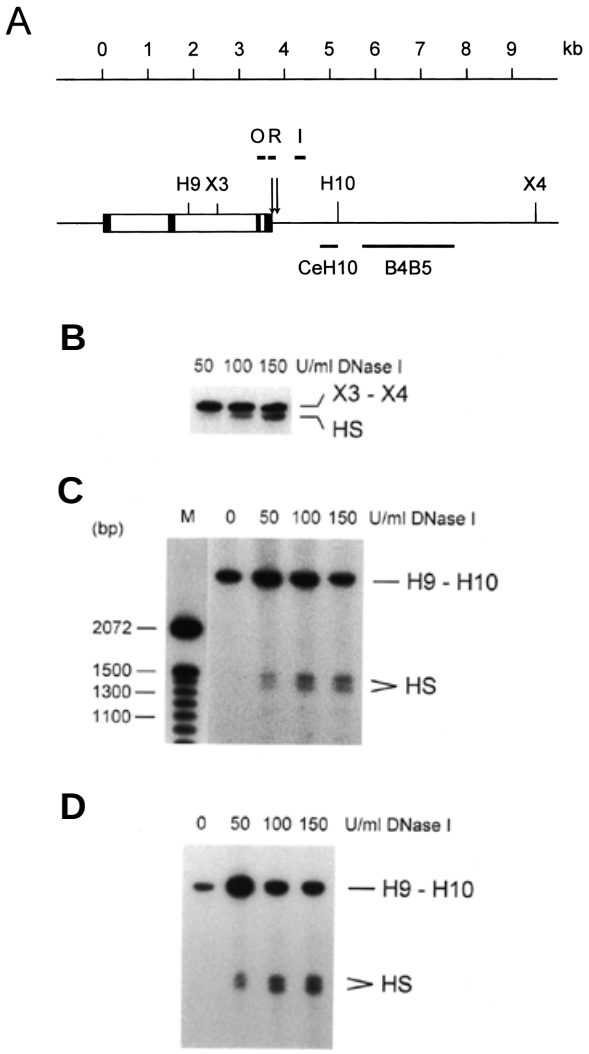

Figure 6. Chromatin structure analysis of the lysozyme OBR. (**A**) Schematic map of the lysozyme gene showing the restriction sites and probes used for indirect end-labelling. X3 and X4 denote *Xba*I sites 3 and 4, H9 and H10 *Hin*dIII sites 9 and 10. Horizontal bars indicate the relative positions of segments O, R and I, and probes CeH10 and B4B5. The DNase I HS sub-sites at +3.9 kb are indicated by vertical arrows. (**B**, **C** and **D**) Mapping of DNase I HS sites in the lysozyme OBR. Nuclei isolated from exponentially growing HD11 (B and C) and DU249 (D) cells were incubated without DNase I or with increasing concentrations of DNase I (50–150 U/ml). Purified DNA samples were then digested with *Xba*I and *Hin*dIII, resolved on 1 and 1.2% agarose gels, respectively, Southern-blotted and hybridised to probes B4B5 and CeH10, respectively. Positions of the genomic fragments X3–X4 and H9–H10 and those additionally cleaved at HS sites (HS) are indicated. Lane M contains a 100 bp DNA ladder.

Yet we have to admit that application of the strategy of using more densely spaced genomic markers might reveal additional weaker OBRs. Indeed, such weaker OBRs likely occur and might have generated, in our previous study, the appearance of a zone of initiation (8). Identification of a second origin (ori-β') within the DHFR initiation zone, ~5 kb downstream from the previously discovered origin, ori-β, is an interesting example of the usefulness of this strategy (4).

A quantitative aspect of the competitive PCR assay is furthermore interesting. Near the center of the OBR (genomic site R) we measured 2500 copies of nascent strands in the 0.8–1.2 kb size fraction (Fig. 2). In a long calculation taking into account the number of cells used (~4×10^8 cells), the amount of nascent DNA for PCR analysis (1.25%), and a cell cycle time of 16 h, and assuming 100% recoveries and that replication forks bi-directionally move at 1.5 kb/min, we estimated the maximal number of copies of nascent 1 kb fragments at the position of

the OBR as ~3600. Thus the number of copies found is very close to the number expected from an OBR.

Our results furthermore show that the lysozyme gene 3' region harbouring the lysozyme OBR contains a CpG island. In the human genome, CpG islands are found at the 5' end of ~56% of the sequenced genes, including all those that are ubiquitously expressed (housekeeping genes) (14,15,39). They are ~1–2 kb in length and mostly include the promoter and one or more of the 5' exons. In spite of the high density of the methylatable sequence CpG, most CpG islands are non-methylated in all tissues, including those in which the gene is silent. The CpG island of the lysozyme gene exhibits an ~66% GC-richness, is ~1.8 kb in length, and contains a high density of CpG doublets (16/100 bp) in its central portion. By use of methylation-sensitive restriction enzymes we show that CpG dinucleotides, that are representatively distributed within the island, and are non-methylated in lysozyme-expressing HD11 cells, non-expressing DU249 cells, and transcriptionally inactive chicken erythrocytes. Interestingly, two CpG sites at the 3' border of the island were found to be methylated in solely HD11 cells, suggesting a function of this methylated site for lysozyme gene expression. Genomic sequencing using sodium bisulfite in order to identify all unmethylated cytosines (40) will be required to complement our view of the methylation status of the lysozyme CpG island. Despite this lack of knowledge it is justifiable to conclude that the 3' end of the lysozyme gene fulfils the characteristics of a CpG island, although it is located at the 3' end of the gene. Only rarely CpG islands have been previously identified at 'non-5' end' locations. Examples are a CpG island located in the second intron of the *Igf2r* gene (41), a CpG island that covers exon 2 of the mouse major histocompatibility complex class II I-Aβ gene (42) and a CpG island found at the 3' region of the glucose-6-phosphate dehydrogenase gene (43). Very recently, it was shown that CpG island-like fragments are enriched in a short nascent strand DNA fraction from erythroleukaemic cells (20). In four cases, three hamster genes and one human gene, the 5' located CpG islands were present in this nascent strand size fraction, but not flanking DNA sequences, suggesting that CpG islands represent a subclass of chromosomal OBRs. Interestingly, the CpG island of the lysozyme gene, though it is located at its 3' end, also possesses an OBR and is, therefore, a member of this subclass.

Besides being localised in a CpG island, the lysozyme OBR is featured by a DNase I HS site that contains two HS subsites spaced by ~100 bp. This poses the question of how the features of the lysozyme gene 3' end discovered here might play a functional role for the action of this site as an origin. It is conceivable that OBRs localise to sites whose chromatin structure allows easy access to replication initiation complexes and auxiliary factors that might regulate the activity of replication origins. Through its location in a non-methylated CpG island the lysozyme OBR should be kept free of methyl-CpG-binding protein 2 (MeCP2/ARBP) and other methyl-CpG binding proteins (44–47). Very recently it was reported that MeCP2/ARBP can recruit a co-repressor complex containing histone deacetylase, which by deacetylation of histones H3 and H4 lysines is capable of inducing or maintaining a repressed chromatin structure (48,49). Non-methylated CpG islands notably exhibit an open chromatin structure and a defined composition (19,50). Specifically, they show a surprisingly high accessibility to restriction enzymes, are deprived of histone H1 and exhibit high levels of acetylation of histones H3 and H4. Furthermore, the discovery of a DNase I HS site in the lysozyme OBR encourages the hypothesis that this feature plays a role in targeting assembly of the replication pre-initiation complex to the OBR. In support of this hypothesis, the DNase I HS site at the lysozyme gene 3' end is a feature of expressing and non-expressing tissues examined (Fig. 6), and thus is likely unrelated to transcription.

Interestingly, the lysozyme gene locus is replicated early during S phase in expressing and non-expressing cell types, i.e. independently of the transcriptional activity of the locus (8). This appears to be an exception, since it is generally assumed that actively transcribed genes are replicated in early S phase, while repressed genes are replicated in late S phase (12,13). Furthermore, we show here that replication of the lysozyme gene locus originates from a non-methylated CpG island. DNA-(cytosine-5) methyltransferase (MCMT) is known to be targeted to replication foci during S phase (51). Recently, Chuang *et al.* showed that MCMT binds to proliferating cell nuclear antigen (PCNA), an auxiliary factor for DNA replication and repair, *in vitro* and *in vivo* (52). The cell cycle regulator p21 negatively regulates targeting of MCMT to PCNA and recruitment of MCMT to the replication machinery, primarily in early S phase. It has been postulated that this reaction protects early replicating CpG islands from methylation. Interestingly, the molecular connections disclosed between cell cycle and non-methylated CpG islands by Chuang *et al.* (52) also provide an explanation for our findings. The methylation-free status of the CpG island at the lysozyme 3' end is provided by the early onset of replication at this site and likely by the inhibition of MCMT by high levels of p21 in early S phase. Through the outlined links, the early onset of replication at the OBR in all tissues, including non-expressing ones (8), is connected with the methylation-free status of the island in all tissues.

ACKNOWLEDGEMENTS

We thank D. Wulf and K. Zimmermann for skilful technical assistance. This work was supported by grants to L.P. from the Deutsche Forschungsgemeinschaft.

REFERENCES

1. DePamphilis,M.L. (1996) In DePamphilis,M.L. (ed.), *DNA Replication in Eukaryotic Cells.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 45–86.
2. Burhans,W.C., Vassilev,L.T., Caddle,M.S., Heintz,N.H. and DePamphilis,M.L. (1990) *Cell*, **62**, 955–965.
3. 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.
4. Kobayashi,T., Rein,T. and DePamphilis,M.L. (1998) *Mol. Cell. Biol.*, **18**, 3266–3277.
5. Yoon,Y., Sanchez,J.A., Brun,C. and Huberman,J.A. (1995) *Mol. Cell. Biol.*, **15**, 2482–2489.
6. Sippel,A.E., Theisen,M., Borgmeyer,U., Strech-Jurk,U., Rupp,R.A.W., Püschel,A.W., Müller,A., Hecht,A., Stief,A. and Grussenmeyer,T. (1988) In Kahn,G. (ed.), *The Architecture of Eukaryotic Genes.* Verlagsgesellschaft Chemie, Weinheim, pp. 355–369.
7. Phi-van,L. and Strätling,W.H. (1988) *EMBO J.*, **7**, 655–664.
8. Phi-van,L., Sellke,C., von Bodenhausen,A. and Strätling,W.H. (1998) *J. Biol. Chem.*, **273**, 18300–18307.
9. Vassilev,L.T. and Johnson,E.M. (1990) *Mol. Cell. Biol.*, **10**, 4899–4904.

10. Kelly,R.E., DeRose,M.L., Draper,B.W. and Wahl,G.M. (1995) *Mol. Cell. Biol.*, **15**, 4136–4148.
11. Gögel,E., Längst,G., Kunkel,E. and Grummt,F. (1996) *Chromosoma*, **104**, 411–518.
12. Goldman,M.A., Holmquist,G.P., Gray,M.C., Caston,L.A. and Nag,A. (1984) *Science*, **224**, 686–692.
13. Holmquist,G.P. (1987) *Am. J. Hum. Genet.*, **40**, 151–173.
14. Antequera,F. and Bird,A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11995–11999.
15. Gardiner-Gardner,M. and Frommer,M. (1987) *J. Mol. Biol.*, **196**, 261–282.
16. Riggs,A.D. and Pfeifer,G.P. (1992) *Trends Genet.*, **8**, 169–174.
17. Barlow,D.P. (1995) *Science*, **270**, 1610–1613.
18. Razin,A. and Cedar,H. (1994) *Cell*, **77**, 473–476.
19. Tazi,J., and Bird,A. (1990) *Cell*, **60**, 909–920.
20. Delgado,S., Gomez,M., Bird,A. and Antequera,F. (1998) *EMBO J.*, **17**, 2426–2435.
21. Cheng,L. and Kelly,T.J. (1989) *Cell*, **59**, 541–551.
22. Simpson,R.T. (1990) *Nature*, **343**, 387–389.
23. Brown,J.A., Holmes,S.G. and Smith,M.M. (1991) *Mol. Cell. Biol.*, **11**, 5301–5311.
24. Jakobovits,E.B., Bratosin,S. and Aloni,Y. (1980) *Nature*, **285**, 263–265.
25. Saragosti,S., Moyne,G. and Yaniv,M. (1980) *Cell*, **20**, 65–73.
26. Diffley,J.F.X., Cocker,J.H., Dowell,S.J. and Rowley,A. (1994) *Cell*, **78**, 303–316.
27. Abdurashidova,G., Riva,S., Biamonti,G., Giacca,M. and Falaschi,A. (1998) *EMBO J.*, **17**, 2961–2969
28. Pemov,A., Bavykin,S. and Hamlin,J. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14757–14762.
29. Beug,H., von Kirchbach,A., Döderlein,G., Conscience,J.-F. and Graf,T. (1979) *Cell*, **18**, 375–390.
30. Langlois,A.J., Lapis,K., Ishizaki,R., Beard,J.W. and Bolognesi,D.P. (1974) *Cancer Res.*, **34**, 1457–1464.
31. Heintz,N.H. and Hamlin,J.L. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 4083–4087.
32. Diviacco,S., Norio,P., Zentilin,L., Menzo,S., Clementi,M., Biamonti,G., Riva,S., Falaschi,A. and Giacca,M (1992) *Gene*, **122**, 313–320.
33. Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
34. Goethe,R. and Phi-van,L. (1997) *J. Leukoc. Biol.*, **61**, 367–374.
35. Hewish,D.R. and Burgoyne,L.A. (1973) *Nucleic Acids Res.*, **52**, 504–510.
36. Wu,C. (1980) *Nature*, **286**, 854–860.
37. Strätling,W.H., Dölle,A. and Sippel,A.E. (1986) *Biochemistry*, **25**, 495–502.
38. Fritton,H.P., Igo-Kemenes,T., Nowock,J., Strech-Jurk,U., Theisen,M. and Sippel,A.E. (1987) *Biol. Chem. Hoppe-Seyler*, **368**, 111–119.
39. Bird,A. (1986) *Nature*, **321**, 209–213.
40. Clark,S.J., Harrison,J., Paul,C.L. and Frommer,M. (1994) *Nucleic Acids Res.*, **22**, 2990–2997.
41. Wutz,A., Smrzka,O.W., Schweifer,N., Schellander,K., Wagner,E.F. and Barlow,D.P. (1997) *Nature*, **389**, 745–749.
42. Macleod,D., Ali,R.R. and Bird,A. (1998) *Mol. Cell. Biol.*, **18**, 4433–4443.
43. Toniolo,D., D'Urso,M., Martini,G., Persico,M., Tufano,V., Battistuzzi,G., and Luzzatto,L. (1984) *EMBO J.*, **3**, 1987–1995.
44. Hendrich,B. and Bird,A. (1998) *Mol. Cell. Biol.*, **18**, 6538–6547.
45. Lewis,J.D., Meehan,R.R., Henzel,W.J., Maurer-Fogy,I., Jeppesen,P., Klein,F. and Bird,A. (1992) *Cell*, **69**, 905–914.
46. Meehan,R.R., Lewis,J.D., McKay,S., Kleiner,E.L. and Bird,A. (1989) *Cell*, **58**, 499–507.
47. Weitzel,J.M., Buhrmester,H., and Strätling,W.H. (1997) *Mol. Cell. Biol.*, **17**, 5656–5666.
48. Jones,P.L., Veenstra,C.J.C., Wade,P.A., Vermaak,D., Kass,S.U., Landsberger,N., Strouboulis,J. and Wolffe,A.P. (1998) *Nature Genet.*, **19**, 187–191.
49. Nan,X., Ng,H.-H., Johnson,C.A., Laherty,C.D., Turner,B.M., Eisenman,R.N. and Bird,A. (1998) *Nature*, **393**, 386–389.
50. Antequera,F., Boyes,J. and Bird,A. (1990) *Cell*, **62**, 503–514.
51. Leonhardt,H., Page,A.W., Weier,H.-U., and Bestor,T.H. (1992) *Cell*, **71**, 865–873.
52. Chuang,L.S.-H., Ian,H.-I., Koh,T.-W., Ng,H.-H., Xu,G., Li,B.F.L. (1997) *Science*, **277**, 1996–2000.