Nuclease-hypersensitive Chromatin Formed by a CpG Island in Human DNA Cloned as an Artificial Chromosome in Yeast*

(Received for publication, October 4, 1999, and in revised form, October 24, 1999)

Mariusz Mucha‡§§, Katarzyna Lisowska‡§, Anna Goc§, and Jan Filipski‡ **

From the §Laboratoire de Mutagénèse, Institut J. Monod, Université Paris VI et Paris VII, 2, place Jussieu Tour 43, 75251 Paris, France and the §Laboratorium Genetyki, Uniwersytet M. Kopernika, ul. Gagarina 11, 87–100 Toruń, Poland

CpG islands are mostly unmethylated GC-, and CpG-rich chromosomal segments overlapping promoter sequences in all housekeeping and many tissue-specific genes in vertebrates. Typically, these islands show an open chromatin structure, low in histone H1 and rich in acetylated histones. We have previously found that the island-like CGCG-rich sites in human DNA are hypersensitive to DNase I upon cloning in Saccharomyces cerevisiae. Here we studied, with a higher resolution, the chromatin formed in yeast by one such site, the CpG island accompanying the human glucose-6-phosphate dehydrogenase gene. We have found two strong hypersensitive sites and several positioned nucleosomes flanking the island despite the absence in yeast of such chromatin fiber-shaping factors as histone H1, methyltransferase, and the tissue-specific transcription factors. This finding, together with similar observations from our laboratories and others supports the idea that variations in GC and/or CpG content substantially contribute to the DNA sequence features modulating the structure of the chromatin. The composition-dependent fluctuations in the accessibility of DNA in the chromatin may constitute an evolutionary advantage and may explain the surprising compositional selection that acts in both the coding and non-coding segments of some genes during mammalian evolution.

CpG islands are GC-rich DNA sequence elements, abundant in unmethylated CpG dinucleotides, that are compositionally clearly distinct from the surrounding DNA. They were identified in the chromosomal segments several years ago (1), but their full functional and structural significance, as well as their evolutionary origin, remains elusive. The islands show variable length (ranging from a few hundred to a few thousand base pairs), variable content of short GC-rich sequence elements (such as the Sp1 transcription factor recognition site), variable position with respect to the start of transcription of neighbor genes, and so on. In addition, it seems that the chromosomal function of these islands varies (2). This variability has made it difficult to discern which features are important and which are secondary for the function of the islands.

One of the most fundamental characteristics of any functionally important segment of DNA within the chromatin fiber is its accessibility to proteins. The simple but efficient test used for screening the chromosome for such sites consists of treatment of the chromatin with DNase I. The chromosomal sites most readily attacked by the enzyme, the hypersensitive sites (HSs) (3), usually correspond to the sites of protein-DNA interactions (3). The CpG islands, as expected for functionally important genomic regions, form special, active, DNase I-hypersensitive chromatin. Histone H1 is present there in low amounts, histones H3 and H4 are acetylated, and nucleosome-free regions are present (4).

Studies of the long-range distribution of hypersensitive sites in the mammalian chromosome showed that the strong HSs are distributed with distances compatible with the size of the chromatin loops (5, 6). We have found a similar pattern of distribution of the HSs in the chromatin of human DNA cloned in yeast artificial chromosomes (YACs) (7). Interestingly enough, the same DNA segments were hypersensitive upon cloning in both organisms. In particular, the locus control region of the β-globin gene cluster, which is nuclease-hypersensitive in human cells, remains hypersensitive upon cloning of the cluster as a YAC (7). Some aspects of the higher order chromatin structure, both structural and functional, are apparently evolutionarily conserved between humans and yeast (7–9). YACs appear to be a simplified but still informative model to study chromatin organization in the chromosome.

Probably the most surprising finding was that several of the HSs in YACs are co-localized with the strong, presumably multiple, CGCG tetranucleotide sites (7). As this tetranucleotide is frequent inside and rare outside the CpG islands, our finding suggested that the CpG islands, which are hypersensitive in vertebrate cells, (4) remain hypersensitive upon cloning in yeast. There are no CpG islands in the yeast genome, and there is a priori no reason why these sites should be hypersensitive in yeast. The work presented here has been performed to have a closer look at the chromatin of a CpG island cloned in yeast. We have chosen the CpG island located at the 3’ end of the human gene encoding glucose-6-phosphate dehydrogenase (G6PD). We have found that the chromatin formed by this island is flanked by two strong hypersensitive sites and is associated with clusters of positioned nucleosomes. This finding confirms our earlier observation (7) and suggests that the evolutionarily conserved GC richness observed in the regulatory regions of eukaryotic genomes (10) might be caused by a selection based...
on an advantage associated with the DNA composition. A possibility should be taken into consideration that GC-rich, as opposed to AT-rich, DNA shows a propensity to form an open chromatin structure.

**MATERIALS AND METHODS**

YAC yWXD206 (alias XY206) cloned by Chen et al. (11) in the AB1380 yeast strain was purchased from ATCC, Manassas, VA (see Fig. 1A for a partial map). It carries the 3' part of the gene coding for G6PD including the first two exons as well as the accompanying CpG island. A DNA segment extracted from the island (nucleotide 2528–2654 from the beginning of the insert) was capable of acting as a core promoter, assuring the regulation of expression of a reporter gene in mammalian but not in Drosophila cells (12). The full nucleotide genomic sequence of the region is available from GenBank™ (Locus HUMFLNG6PD). The integrity of the YAC was verified by mapping with Mun1, Xba1, BssHIII, and SfiI restriction enzymes. The hybridizations of the products of partial digestion of the YAC with these enzymes were performed using radioactive C and T probes. Probe T was a BamHI–PvuII fragment and probe C a PvuII–PstI fragment of pBR322 recognizing the YAC telomeres t and c (distal and proximal to the centromere, respectively). The YAC did not show any signs of major rearrangements. The probe G6PD7 used in the mapping of the positions of nucleosomes was prepared by a polymerase chain reaction using primers 5'-CACACACCGGCTCTCCTCGT and 5'-GCAAGAATCTCTATCGACCAGTGGG.

The micrococcal nuclease digestion was performed as described (13) using the nuclei isolated as described previously (7) and suspended in MN buffer (10 mM Tris, pH 7.5, 60 mM KCl, 1 mM CaCl2, 15 mM NaCl). The DNA isolated from the reaction mixture was subsequently cut by the SfiI restriction enzyme and analyzed by electrophoresis in a 0.5% agarose gel and hybridization with the G6PD7 probe radioactively labeled as described before (7). The DNase I digestion performed to localize the HS has been described before (7).

**RESULTS AND DISCUSSION**

Among the chromosomal sites hypersensitive to DNase I in YACs carrying human DNA, some are particularly sensitive to the restriction enzyme that recognizes CGCG tetranucleotides (7). This observation suggested that CpG islands, which are nuclease-hypersensitive in humans, remain hypersensitive upon cloning in yeast. To verify this hypothesis we have chosen the YAC carrying part of the gene coding for glucose-6-phosphate dehydrogenase, including its accompanying CpG island.

Foremost, we wanted to ascertain that the DNA of the chosen YAC had not been rearranged. For that reason we mapped the MunI and XbaI (not shown) as well as the BssHIII and SfiI restriction sites (Fig. 1A) in the YAC. Their relative positions were concordant with those predicted on the basis of the nucleotide sequence extracted from GenBank™. In addition, the restriction map showed that the CpG island was located close to the telomere proximal to the centromere and that the YAC was somewhat shorter (31.7 kb) than reported in the literature (40 kb) (14).

The mapping of the site hypersensitive to DNase I in the YAC has been performed by treating the yeast chromatin with various concentrations of the enzyme. The resulting DNA products of partial digestion were size-fractionated by gel electrophoresis, transferred to a nylon membrane, and revealed by hybridization with the radioactively labeled telomeric probe C. Fig. 1B (lanes 2–4) shows the result of hybridization of this DNA along with the hybridization of the product of partial digestion with the intact DNA with the BssHIII restriction enzyme (lane 1). The latter enzyme recognizes the CGCGCG hexanucleotide, three copies of which (two of them overlapping) are clustered in the island, that generates a single band in the autoradiogram in the region of the island (Fig. 1B, lane 1). Comparison of the positions of the DNase I bands with that of the BssHIII band reveals that two strong hypersensitive sites (Fig. 1B) flank the CpG island in this YAC. The DNase I hypersensitivity depends on the chromatin structure, because the digestion of naked DNA with this enzyme did not produce any strong band on the autoradiogram in the corresponding position (not shown).

The HS, including those formed by the CpG islands, are usually accompanied by positioned nucleosomes (4). This also appeared to be the case for the island studied here. The digestion of the YAC chromatin with micrococcal nuclease produces a ladder of bands compatible with the distances between positioned nucleosomes in the 6742-base pair-long SfiI segment containing the island (see Fig. 2). The positioning extends into the transcriptional unit, about 1 kb beyond the limits of the CpG island. This may suggest that this region contributes to the chromatin organization of the G6PD gene in human cells and to the control of its expression (12).

Several factors involved in the modulaion of the accessibility of DNA along the chromatin fiber in higher eukaryotes are absent from yeast. First, the yeast lack CpG methylation activity. Second, yeast also lack the standard histone H1 (15). Finally, formation of hypersensitive sites and an active chromatin accompanied by nucleosome positioning and histone acetylation is believed to be associated with the formation of complexes between transcription factors and their respective recognition sites in the DNA. Some of the yeast transcription factors such as MIG1 or RAP1 apparently recognize the same
segments of yeast DNA extracted from different parts of the perspective sites appear even at the boundaries between two tions were inserted into the yeast natural chromosomes. Hy-
DNA junctions if the artificial constructs containing such junc-
tsensitive sites in YACs (see Figs. 4 and 5 in Ref. 7).
This finding confirms and extends our previous results, show-
occupies its correct place in the region of CpG island (Fig. 1).
despite that difference, the yeast cell is able to organize the tion of yeast and mammalian promoters (18). Surprisingly,
demonstrated that there is a basic difference between the organisa-
tion (16) and, more importantly, it has been demon-
strated, and frequently recombining (28) chromosomal seg-
menting during DNA synthesis during replication (26) or repair (27) of the early replicating segments of the chromosomes might have filled these early replicating (26), actively transcribed, and frequently recombining (28) chromosomal segments with G and C. The late replicating regions might be under AT pressure like the genomes of some bacterial species (29). The selectionists suggested that the process of evolution-
ary selection is involved in the regional enrichment of genomes in GC base pairs based on their higher thermal stability com-
ception sequences as mammalian transcription factors (16, 17). The recognition sequence of MIG1 (CGCGGG) is even present in multiple copies in the CpG island studied. However, it has been shown that the presence of the consensus sequence alone is necessary but not sufficient for the DNA-MIG1 complex formation (16) and, more importantly, it has been demonstrated that there is a basic difference between the organization of yeast and mammalian promoters (18). Surprisingly, despite that difference, the yeast cell is able to organize the chromatin in the YAC in such a way that the open chromatin occupies its correct place in the region of CpG island (Fig. 1). This finding confirms and extends our previous results, showing co-localization of CGCG sites with the DNase I-hypersensitive sites in YACs (see Figs. 4 and 5 in Ref. 7).

Some earlier experiments from other laboratories (19) showed that pairs of HS, accompanied by arrays of positioned nucleosomes, systematically appeared at the pBR322-yeast DNA junctions if the artificial constructs containing such junctions were inserted into the yeast natural chromosomes. Hypersensitive sites appear even at the boundaries between two segments of yeast DNA extracted from different parts of the yeast genome (20–22). The authors of these observations sug-
gested that hypersensitivity arose either as a result of disruption of the arrangement of sequences that direct assembly of nucleosomes into regular arrays or by creating a collision between phased nucleosomes directed by different sequences (22).

We suggest that the appearance of these boundaries is related to the changes in the GC content and/or frequency of CpG dinucleotides. Such a change occurs at the extremities of a CpG island as well as at the borders of segments of the CpG-rich plasmid pBR322 embedded into yeast DNA. Probably, an uniden-
ified factor(s) that is dependent on the local GC content and/or density of CpG dinucleotides in the DNA molecule influences the structure and/or the mobility of nucleosomes, and as a consequence, influences the structure of the chromatin fiber. Alternatively, the GC-rich DNA may have an intrinsic tendency to form a different kind of chromatin fiber than the AT-rich DNA. The border between the two types of fiber generates a structural discontinuity having the properties of a hypersensitive site.

Three independent lines of evidence suggest that a GC and/or CpG frequency-related feature has an impact on the structure and function of the chromatin. First, there is the sensitivity of the chromatin formed in yeast by a heterologous CGCG-(7) and GC-rich (19) DNA to DNase I. A second line of evidence comes from an analysis of CpG distribution in DNA. It appears (10) that the observed versus expected density of CpG dinucleotides is elevated in a 1–2-kb-long segment overlapping the regulatory region of an average eukaryotic gene. This rule holds for all groups of eukaryotic organisms (as opposed to bacteria) irrespective of whether the organism methylates its CpG (like humans) or not (like yeast or Drosophila). We suppose that the high CpG content in these regions has evolved as the result of a requirement for nucleosomal positioning and for protein-accessible chromatin structure in these regions. The third piece of evidence has been obtained by population genetic tests. The analysis of the distribution of alleles of major histocompatibility complex genes in several mammalian species has demonstrated the existence of GC content-driven evolutionary selection in both coding and noncoding regions of these genes, including the CpG islands, introns, and silent positions in the coding regions (23). In other words, the carriers of an allele of a major histocompatibility complex gene in which an AT base pair has been replaced by a GC base pair produce more abundant progeny than the carriers of the wild type allele, even if the mutation has not changed the protein coded by these genes.

This result contributes to the old but still alive discussion (23, 24) between neutralists and selectionists concerning the origin of the compositional compartments observed in some genomes. For example, in the human genome the housekeeping genes and their accompanying non-coding sequences are GC-rich (25), whereas the tissue-specific genes are either GC- or AT-rich. Neutralists suggested that GC-biased mutations occurring during DNA synthesis during replication (26) or repair (27) of the early replicating segments of the chromosomes might have filled these early replicating (26), actively transcribed, and frequently recombining (28) chromosomal segments with G and C. The late replicating regions might be under AT pressure like the genomes of some bacterial species (29). The selectionists suggested that the process of evolution-
ary selection is involved in the regional enrichment of genomes in GC base pairs based on their higher thermal stability com-
ference nucleosome. After extraction, the DNA was digested with the restriction enzyme SfiI, size fractionated on a 0.5% agarose gel, transferred to a nylon filter, and hybridized to the radioactively labeled probe G6PD7.

A

B

FIG. 2. Micrococcal nuclease (Mnase)-sensitive sites accompa-
yning the CpG island in the XY206 YAC. A, the result of hybridiza-
tion of the product of partial digestion of the yeast chromatin with Mnase. After extraction, the DNA was digested with the restriction enzyme SfiI, size fractionated on a 0.5% agarose gel, transferred to a nylon filter, and hybridized to the radioactively labeled probe G6PD7.

The numbers on the left show the positions of the Mnase-sensitive sites with respect to Nt. 1 (see Fig. 1). Lanes 1 and 2 represent the product of partial digestion of the yeast chromatin with 5 and 10 units/ml Mnase, respectively, and shows the same area of the filter as lanes 3 and 4, the latter being exposed longer to visualize the faint bands. B, interpretation of the result of the mapping. Circles represent the nu-
cleosomes, and thick arrows represent the hypersensitive sites deter-
mined in the experiment shown in Fig. 1. Thin arrows represent the position of the restriction enzyme SfiI and the limits of the exons. The dark box represents the CpG island, and the dashed lines represent the position of the Mnase-sensitive sites. The white box ("G6PD7") represents the position of the probe used in the indirect end-labeling experiment.

| Lane | Position |
|------|----------|
| 1    | 1.93     |
| 2    | 2.97     |
| 3    | 3.98     |
| 4    | 4.08     |

| Lane | Position |
|------|----------|
| 1    | 3.82     |
| 2    | 4.08     |
| 3    | 5.15     |
| 4    | 6.15     |

| Lane | Position |
|------|----------|
| 1    | 7.55     |
| 2    | 8.77     |
| 3    | 8.58     |
| 4    | 7.77     |

| Lane | Position |
|------|----------|
| 1    | 1.54     |
| 2    | 2.13     |
| 3    | 2.79     |
| 4    | 3.18     |

| Lane | Position |
|------|----------|
| 1    | 3.02     |
| 2    | 3.51     |
| 3    | 4.06     |
| 4    | 4.51     |
pared with AT base pairs (30). As this explanation does not seem to be particularly convincing, the demonstration that composition-driven selection indeed takes place came as a surprise (23).

The results presented here, together with the results from other laboratories mentioned under “Results and Discussion” (10, 19, 23) allow us to suggest a new molecular base to explain the observed compositional selection in the crucial regions of the vertebrate chromosome. The DNA with high GC content and/or a high proportion of CpG dinucleotides (which is proportional to the square of the GC content) might have an intrinsic tendency to form an active chromatin structure. This may facilitate interaction with regulatory proteins in the region of promoters, simplifying the regulation of transcription. The easy formation of an active chromatin may also assure the smooth passage of the chromatin fiber carrying the transcription unit through the transcription factories (31). This advantage might be high enough in the case of some important proteins to become a base for evolutionary selection of GC base pairs. It might also be the reason why some organisms have chosen, during evolution, GC-biased repair processes to repair their early replicating or frequently recombining genes in the germline. The combined effect of these processes may explain such a strange observation as the difference between functional germline. The combined effect of these processes may explain such a strange observation as the difference between functional germline.

This hypothesis is in agreement with the contemporary consensus opinion on the formation of chromatin structure. The current model states that methylation, tissue-specific transcription factor binding, acetylation, and histone H1 binding stabilize and provide fine tuning to the chromatin accessibility pattern. Contributing to that are other DNA features and trans-acting factors that also affect the chromatin (33–35). The evidence presented here suggests that one of these features could be the high GC content and/or concentration of CpG dinucleotides in the DNA sequence, which might contribute to the organization of the open chromatin structure near eukaryotic genes.