A Distinct Sequence (ATAAA)_n Separates Methylated and Unmethylated Domains at the 5′-End of the GSTP1 CpG Island

Received for publication, August 16, 1999, and in revised form, April 19, 2000
Published, JBC Papers in Press, April 21, 2000, DOI 10.1074/jbc.M906538199

Douglas S. Millar‡, Cheryl L. Paul‡, Peter L. Molloy§, and Susan J. Clark‡§

From the ‡Kanematsu Laboratories, Royal Prince Alfred Hospital, Missenden Road, Camperdown, New South Wales 2050, Australia, §Commonwealth Scientific and Industrial Research Organisation Molecular Science, P. O. Box 184, North Ryde, New South Wales 1670, Australia, and the ¶Department of Medicine, University of Sydney, Sydney, New South Wales 2050, Australia

What defines the boundaries between methylated and unmethylated domains in the genome is unclear. In this study we used bisulfite genomic sequencing to map the boundaries of methylation that flank the 5′- and 3′-ends of the CpG island spanning the promoter region of the glutathione S-transferase (GSTP1) gene. We show that GSTP1 is expressed in a wide range of tissues including brain, lung, skeletal muscle, spleen, pancreas, bone marrow, prostate, heart, and blood and that this expression is associated with the CpG island being unmethylated. In these normal tissues a marked boundary was found to separate the methylated and unmethylated regions of the gene at the 5′-flank of the CpG island, and this boundary correlated with an (ATAAA)_19–24 repeated sequence. In contrast, the 3′-end of the CpG island was not marked by a sharp transition in methylation but by a gradual change in methylation density over about 500 base pairs. In normal tissue the sequences on either side of the 5′-boundary appear to lie in separate domains in which CpG methylation is independently controlled. These separate methylation domains are lost in all prostate cancer where GSTP1 expression is silenced and methylation extends throughout the island and spans across both the 5′- and 3′-boundary regions.

Almost all organisms with a genome of greater than 10^9 base pairs methylate their DNA, generally at the 5′-position of cytosine in CpG dinucleotides. In vertebrate genomes CpG sites occur about one-fifth of that expected on the basis of base composition. This has been attributed to the inherent mutability of 5-methyl cytosine to thymine (1). However, CpG-rich clusters, termed CpG islands, are found interspersed in large regions of the CpG-depleted sequence (2). The maintenance of these CpG islands is indicative of a functional role, and it is nearly always maintained in an unmethylated state (5). Methylation of CpG islands can occur, however, for example on the inactive X chromosome, in promoters of imprinted genes and in aberrant methylation associated with oncogenesis (6). In all of these cases methylation of CpG islands spanning the gene promoter regions is strongly associated with transcriptional silencing (7).

As yet there is no clear understanding of what permits CpG islands to be preferentially unmethylated, what defines the boundary of the unmethylated region, and what leads to the aberrant methylation of CpG islands commonly seen in cancer. It has been suggested that maintenance of the unmethylated CpG islands is dependent on continued active transcription and/or on the binding of specific proteins that may protect them from methylation (8). Analysis of the CpG island of the adenine phosphoribosyl transferase gene in transgenic mice, using mutated promoters, has shown that Sp1 sites are important for maintaining this CpG island in its unmethylated state (9, 10). Thus loss of specific proteins or interference with their binding or loss of active transcription may all contribute to CpG island promoter methylation in cancer.

The GSTP1 gene that encodes glutathione S-transferase-π (GST-π) contains a typical CpG island promoter region and is widely expressed in most tissues, with the exception of liver (11). It has recently been demonstrated that GSTP1 becomes methylated in a high proportion of prostate cancers and that this methylation is accompanied by gene silencing (12, 13, 14). Detailed bisulfite sequencing analysis of the CpG island spanning the core promoter region of the GSTP1 gene showed that methylation is extensive at essentially all CpG sites in prostate cancer DNA (14). In contrast, the CpG island is completely unmethylated in normal prostate tissue. However methylation was found to occur outside the CpG island in the CpG-depleted 3′-region of the GSTP1 gene in both normal and cancer tissue (14).

In this paper we have extended the bisulfite sequence analysis of the methylation profile of the GSTP1 CpG island/promoter to include the flanking regions where the normal and prostate cancer methylation profiles diverge. This has allowed identification of a defined “boundary” region of methylation at the 5′-flank, which is marked by a repeating ATAAA sequence.

MATERIALS AND METHODS

DNA Samples—DNA from three primary prostate cancer samples and corresponding normal DNA from the diseased prostate were isolated as follows. Tissue samples were isolated from patients undergoing radical prostatectomy for prostate cancer. Tissue slices, about 4-mm
thick, were snap frozen in liquid nitrogen, and histology was performed on adjacent slices to identify regions of tumor and normal tissue. Tissue samples were isolated from the frozen slice using a punch and ground into a powder under liquid nitrogen using a mortar and pestle. DNA was isolated using Trizol® reagent (Life Technologies, Inc.) according to the manufacturer’s protocols. DNA was further treated with RNase A and then proteinase K before phenol extraction and ethanol precipitation. DNA from the prostate tumor cell lines LNCaP and PC3 was prepared as described previously (14). DNA from the following normal tissues were isolated from the frozen slice using a punch and ground into a powder under liquid nitrogen using a mortar and pestle. DNA samples were isolated from the frozen slice using a punch and ground into a powder under liquid nitrogen using a mortar and pestle.

**Bisulfite Conversion**—Sodium bisulfite converts cytosine residues to uracil residues in single-stranded DNA under conditions whereby 5-methylcytosine remains nonreactive. All cytosine residues remaining in the target sequence after PCR amplification represent previously methylated cytosines. The bisulfite reaction was carried out on 1–2 μg of HindIII-digested patient DNA for 16 h at 55 °C under conditions described by Clark et al. (15) and Clark and Frommer (16). The samples were purified using Wizard DNA Clean-Up System desalting columns (Promega), eluted in 50 μl of H2O, and incubated with 5 μl of 3 M NaOH for 15 min at 37 °C. The solutions were neutralized by the addition of NH4OAc, pH 7, to 3 M, and the DNA was ethanol precipitated, dried, resuspended in 10 μl of 10 mM Tris-HCl (pH 8), 0.1 mM EDTA in the case of tissue samples and 50 μl for cell line DNA and stored at −20 °C. PCR amplification reactions were performed within 24 h of bisulfite conversion.

**PCR Amplification and Primers**—PCR amplifications were performed in 50-μl reaction mixtures containing 2 μl of bisulfite-treated genomic DNA, 200 μM of each of the four dNTPs, 6 ng/μl of each of the primers, 1–2 mM MgCl2, 2 units of AmpliTaq DNA polymerase (PerkinElmer), and reaction buffer consisting of 67 mM Tris, 16.6 mM ammonium sulfate, 1.7 mM MgCl2 bovine serum albumin, and 10 mM β-mercaptoethanol in buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA). The strand-specific nested primers used for amplification of bisulfite-treated DNA are indicated in Table I.

**Sequence Analysis**—The PCR fragments (1-IV) were cloned using the pCR-Script™ Amp SK(+) cloning kit (Stratagene) according to the manufacturers instructions. Individual clones were either sequenced manually using Sequenase version 2.0 DNA sequencing kit (USB) or using PRISM™ DyeDeoxy Terminator Cycle Sequencing Kit (PE/ABI) with AmpliTaq® DNA polymerase and the automated 373A DNA Sequencer (ABI). Details of the Genescan analysis are described by Millar et al. (14). The percent methylation is calculated using peak height of C versus peak height of T for each position.

**Northern Blot Analysis**—Multiple tissue Northern blots I and II were purchased from CLONTECH and probed with a PCR fragment generated (see Table I) from exon 7 of the GSTP1 transcript according to the manufacturer’s instructions. The blots were then stripped and reprobed with the β-actin internal probe supplied to determine the loading levels of each mRNA species. Blots were exposed and quantified using a Molecular Dynamics PhosphorImager and ImageQuant software.

## RESULTS

Like many “housekeeping” genes the GSTP1 gene contains a typical CpG island that extends from −400 base pairs upstream to 800 base pairs downstream of the transcription initiation site (Fig. 1A). We have previously shown that the core promoter region is unmethylated in normal prostate cells but becomes methylated in prostate cancer. To understand the mechanism responsible for abnormal methylation of this CpG island in prostate cancer tissues we have now: 1) extended the methylation analysis of the GSTP1 gene across the entire CpG island in a range of normal tissues, particularly in relation to the extent of the unmethylated domain and the nature of its boundaries and 2) compared the methylation profile of the GSTP1 gene in these domains to the methylation profile in prostate cancer cells.

**Expression and Methylation Profile of the GSTP1 Promoter in Normal Tissues**—The GST-π protein is known to be widely expressed in most tissues (11), but it is not known if the methylation pattern is equivalent in these tissues. Therefore, we examined the expression and methylation profile of the GSTP1 gene in a number of normal tissues including heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, and peripheral blood. Fig. 2 shows the tissue distribution of expression of GSTP1 by hybridization to multiple tissue cDNA Northern blots. Extensive expression was seen in almost all tissues in agreement with data obtained from immunohistochmical analysis and studies of GST enzyme expression (11). However, lowered GSTP1 expression was observed in the liver; this is also consistent with previous data on protein levels (17, 18).

We first examined the methylation profile of the CpG-rich
Identification of a Methylation Boundary in GSTP1

promoter and upstream flanking region. This region spanned 66 CpG sites (−56 to +30) corresponding to the coordinates base numbers from 520 to 957 (GenBank accessions M24485) was amplified from normal prostate tissue and cloned. The methylation profile of 15 individual clones is shown. Closed circles indicate a methylated CpG and open circles indicate an unmethylated CpG. The number of CpGs at any one site is shown. The degree of methylation is indicated by the + number, where − is 0%, + is 1–25%, ++ is 26–50%, +++ is 51–75%, and ++++ is 76–100%. polymorphisms and therefore could not be used reliably for direct sequence analysis; consequently levels of methylation for the 5′-upstream region were determined by cloning the PCR product and sequencing individual clones. The polymorphisms found in the 5′-upstream region include variation in the number and fidelity of (ATATA) repeats as well as sequence polymorphisms that include CpG site polymorphisms. An example of the methylation profile obtained for the clones generated from the PCR fragment amplified from normal prostate DNA is shown in Fig. 3. As can be seen the methylation profile is heterogenous; in fact a heterogenous methylation profile was common to all the normal tissue samples tested. The methylation data generated from the clonal analysis was averaged for each CpG site and as such corresponds to the direct PCR sequencing analysis.

Analysis of the extent of methylation at individual CpG sites in the normal tissues, prostate, blood, brain, spleen, smooth muscle, lung, bone marrow, pancreas, and heart showed no methylation at all in the core promoter region (CpGs −28 to +10) (Fig. 4B). The exception was in normal liver DNA where significant methylation was seen in a cluster of sites, from CpG site −7 to CpG site +7 encompassing the transcription start site. A similar methylation profile was observed in two other samples of normal liver DNA that were sequenced (data not shown). Interestingly, liver was the only normal tissue examined that showed reduced GSTP1 expression, indicating a correlation between the methylation of these sites and expression.

In the 5′-upstream region (CpGs −56 to +30), a marked change was observed in the methylation pattern of all the normal tissues examined (Fig. 4A). There was essentially no methylation in all the normal tissues up to and including CpG site −43; in contrast further upstream, from CpG site −44 and beyond, there was extensive methylation at most CpG sites. However, considerable variability was noted between different tissues in the level and pattern of methylation. Some sites (CpG sites −56, −55, and −53) were heavily methylated (75–100%) in nearly all tissues, whereas CpG site −52 and −48 was notable in that it was commonly unmethylated or undermethylated in some tissues. Also the six CpG sites, −44 to −49, were not methylated in smooth muscle cell DNA. Interestingly the abrupt transition from the DNA domain that is extensively methylated in all normal tissues to that which is unmethylated.
ated, between CpG sites -44 and -43, corresponds to the location of an (ATAAA)_{19-24} repeat sequence (Fig. 1C). Moreover, this pattern of methylation was found to occur in both GSTP1 alleles (data not shown). Immediately upstream of the (ATAAA)_{19-24} repeat are two members of the Alu interspersed repeat sequence family (Fig. 1). The CpG sites (spanning CpGs -56 to -44) in the Alu sequence immediately adjacent to the (ATAAA)_{19-24} repeat were hypermethylated in all normal tissues examined (Fig. 4A). The (ATAAA)_{19-24} repeat thus corresponds to a distinct boundary that separates the methylated Alu repeat DNA at the 5'-flank of the CpG island from the unmethylated DNA within the CpG island.

In contrast there is no clear sequence boundary at the 3'-end of the GSTP1 CpG island. Moreover the 3'-flank of the CpG island does not contain any Alu repeats or any other obvious repeat sequences. We previously noted that in normal prostate tissue the CpG island, up to CpG site 33 as well as sites 52 and 53, was unmethylated whereas the 3'-end of the GSTP1 gene from CpG site 69 to 103 was extensively methylated (14). To locate the exact methylation boundary at the 3'-end of CpG island, we have now sequenced a further 1 kilobase, from CpG site 13 in intron 1 to CpG site 67 in intron 4 (PCRIII and -IV, Fig. 1), from four normal prostate tissue samples. As shown in Fig. 5, there is no methylation in the CpG sites within the CpG island; however, there is considerable heterogeneity in the methylation profile at the 3'-flank of the CpG island and more CpG sites become methylated further downstream from the CpG island region. Interestingly, the first site that is methylated is CpG 31, which is located at the junction of intron 1 and exon 2. The profile of each of the four normal prostate samples is

![Methylation analysis of the 5'-CpG island flank of the GSTP1 gene in normal tissues.](image)

FIG. 4. Methylation analysis of the 5'-CpG island flank of the GSTP1 gene in normal tissues. A, represents the combined methylation profile of individual clones from the 5'-upstream PCR region (PCRI). The number of individual clones sequenced from each tissue type is indicated at the top of the panel. The base number relative to the GSTP1 sequence (GenBank™ M24485) and the CpG site numbers relative to the start of transcription are indicated. The double line represents the position of the (ATAAA)_{19-24} repeat sequence. Quantitation of DNA methylation levels are determined by averaging the methylation state of each CpG site from each set of clones, as described in Fig. 3. The average methylation levels are indicated by: 0%, 0–25%; 1, 26–50%; 2, 51–75%; and 3, 76–100%. B, represents the direct sequence analysis on PCR products obtained from the core promoter region (PCRII) of the same tissue samples analyzed in A. DNA methylation levels were determined by quantitative genescan analysis as described by Millar et al. (14). The average methylation levels are indicated by: 0%, 0–25%; 1, 26–50%; 2, 51–75%; and 3, 76–100%. a indicates a CpG site where the methylation was not determined.
Identification of a Methylation Boundary in GSTP1

**Table 1**

| Base Number | CpG Site | CN n=6 | NN n=8 | BN n=6 | GN n=12 |
|-------------|----------|--------|--------|--------|--------|
| 1338        | 13       | -      | -      | -      | -      |
| 1351        | 14       | -      | -      | -      | -      |
| 1360        | 15       | -      | -      | -      | -      |
| 1367        | 16       | -      | -      | -      | -      |
| 1372        | 17       | -      | -      | -      | -      |
| 1394        | 18       | -      | -      | -      | -      |
| 1406        | 19       | -      | -      | -      | -      |
| 1412        | 20       | -      | -      | -      | -      |
| 1428        | 21       | -      | -      | -      | -      |
| 1433        | 22       | -      | -      | -      | -      |
| 1438        | 23       | -      | -      | -      | -      |
| 1456        | 24       | -      | -      | -      | -      |
| 1462        | 25       | -      | -      | -      | -      |
| 1482        | 26       | -      | -      | -      | -      |
| 1489        | 27       | -      | -      | -      | -      |
| 1491        | 28       | -      | -      | -      | -      |
| 1509        | 29       | -      | -      | -      | -      |
| 1535        | 30       | -      | -      | -      | -      |
| 1546        | 31       | ++     | -      | -      | -      |
| 1556        | 32       | -      | -      | ++     | -      |
| 1575        | 33       | -      | -      | -      | ++     |
| 1640        | 34       | -      | -      | -      | -      |
| 1648        | 35       | -      | -      | -      | -      |
| 1658        | 36       | +      | ++++   | -      | -      |
| 1676        | 37       | -      | -      | -      | -      |
| 1679        | 38       | -      | -      | -      | -      |
| 1695        | 39       | -      | -      | -      | ++     |
| 1715        | 40       | ++     | -      | -      | -      |
| 1724        | 41       | -      | -      | -      | -      |
| 1740        | 42       | -      | -      | -      | -      |
| 1764        | 43       | -      | -      | +      | -      |
| 1812        | 44       | -      | -      | -      | -      |
| 1822        | 45       | -      | -      | -      | -      |
| 1827        | 46       | -      | +      | -      | +      |
| 1833        | 47       | -      | -      | -      | -      |
| 1838        | 48       | -      | -      | -      | -      |
| 1852        | 49       | +      | -      | -      | -      |
| 1868        | 50       | +      | -      | -      | -      |
| 1875        | 51       | +      | -      | -      | -      |
| 1880        | 52       | -      | -      | +      | ++++   |
| 1882        | 53       | ++     | -      | -      | -      |
| 1890        | 54       | +      | +      | ++     | ++++   |
| 2042        | 59       | ++++   | +      | +      | ++++   |
| 2069        | 60       | ++++   | +      | -      | ++++   |
| 2099        | 61       | +      | -      | +      | ++++   |
| 2123        | 62       | +      | +      | ++++   | ++++   |
| 2160        | 63       | ++++   | -      | +      | ++++   |
| 2172        | 64       | ++++   | -      | +      | ++++   |
| 2233        | 65       | +      | +      | ++++   | ++++   |
| 2262        | 66       | +++    | ++     | ++++   | ++++   |
| 2305        | 67       | ++++   | ++++   | ++++   | ++++   |

**Fig. 5.** Methylation analysis of the 3’-CpG island flank of the GSTP1 gene in normal prostate tissues. The methylation data represent the combined methylation profile of individual clones from two 3’-downstream PCR regions (CpG 13–54, PCRIII) and (CpG 59–67, PCRIV). The number of individual clones sequenced from each tissue type is indicated at the top of the panel. The base number relative to the GSTP1 sequence (GenBank™ M24485) and the CpG site numbers relative to the start of transcription are indicated. DNA methylation levels are indicated as in Fig. 4A.

**Discussion**

It is well established that CpG islands remain unmethylated in normal cells and are generally associated with transcriptionally active genes (5). However, in cancer hypermethylation of CpG islands is a common aberration and this is often associated with gene silencing (6). What has not yet been established is what defines the unmethylated CpG island domain, that is are there distinct sequence boundaries between methylated and unmethylated molecules.

We have previously shown that the core promoter region (CpGs −28 to +10) and all sites examined in the body and 3’-end of the gene are extensively methylated in prostate cancer DNA (14). We have now extended this analysis to examine the extent of methylation in the region upstream of the core promoter in the region (CpGs −56 to −30) where we have found a marked boundary of methylation in normal tissue. We examined the methylation state in two prostate cancer cell lines, PC-3, which expresses GSTP1 and LNCaP, a GSTP1 nonexpressing cell line. Using reverse transcriptase-PCR we have shown that the treatment of LNCaP cells with 5-azacytidine, a demethylating agent, reactivates GSTP1 mRNA expression (data not shown). In addition, we analyzed prostate tumor DNA from three patients (BC, CC, and DC) and DNA isolated from a histologically normal region of the prostate of one of these patients (BN).

For LNCaP cells that do not express GSTP1 the extensive high level of methylation seen in the core promoter region was found to continue through the upstream region to and beyond the ATAAA boundary (Fig. 6A). The methylation profile of DNA from PC3 cells that express GSTP1 was distinctly different from that of LNCaP cells. There was very little methylation in the upstream region except beyond the boundary region from (CpG −46 to −56). The low methylation observed is consistent with the generally lower methylation level seen in the core promoter (Fig. 6B). Immunohistochemical analysis of both cell lines (data not shown) demonstrated that in the fully methylated LNCaP cells the GST-π protein was not expressed. PC-3 cells on the other hand harbored two distinct classes of cells. The majority of the cells expressed abundant GST-π protein, whereas in a subset of cells the GST-π protein could not be detected. This is consistent with the methylation patterns observed, that is the patterns represent an average of the methylation profile of the pooled DNA that contain a mixture of methylated and unmethylated molecules.

In all three primary prostate cancer specimens (BC, CC, and DC), extensive methylation was seen downstream of the ATAAA boundary (−43 CpG to +10 CpG) (Fig. 6, A and B). The methylation pattern was heterogeneous at any one CpG site and varied between patient samples. The levels of methylation across the promoter region varied from 25 to 75% methylation for each CpG site, as determined by direct PCR sequencing (Fig. 6B). However, when we examined clones derived from the core promoter PCRII fragment we found that two distinct classes of molecules existed, those that were completely unmethylated and clones that showed extensive methylation through this region (data not shown). Because the cancer samples are not homogeneous and contain a mixture of normal and cancerous prostate epithelial cells, as well as stromal elements, it is likely that the unmethylated clones derive from contaminating normal cells. Therefore in the analysis of the clonal data from PCR1, presented in Fig. 6A, we have only presented the frequency of methylation of individual CpG sites among the population of methylated molecules. This explains why the methylation pattern in the cancer samples appears to be more intense from CpG −43 to −30 than from CpG −28 to CpG +10. As we have reported for GSTP1 and other genes methylated in cancer (14, 19, 20), individual differences in methylation of specific CpG sites, e.g. CpG site −56, were evident between the different patients. Beyond the boundary region (CpG −43), the extent and pattern of methylation was essentially similar in the three cancer samples and in the two normal prostate DNA samples studied, normal prostate and BN (Fig. 4A and 6A).

**DISCUSSION**

Different but all lack a distinct methylation boundary gradually increasing in methylation density over a 500-base pair region (see Fig. 7).

**Methylation Profile in Prostate Cancer DNA**—We have previously shown that the core promoter region (CpGs −28 to +10) and all sites examined in the body and 3’-end of the gene are extensively methylated in prostate cancer DNA (14). We have now extended this analysis to examine the extent of methylation...
and unmethylated sequences or do the methylation profiles merge, and moreover how are these boundaries disrupted or bypassed in a cancer cell. To define the sequence boundaries between the methylated and unmethylated domains of a CpG island, we have finely mapped the methylation profile of the upstream and downstream ends of the GSTP1 CpG island in a number of normal tissues including prostate tissue. The GSTP1 gene was chosen because it has a typical CpG island spanning the promoter and exons 1–3 of the gene, it is known to be expressed in a range of normal tissues, and it is frequently hypermethylated in prostate cancer.

A schematic representation of the methylation profiles across the GSTP1 CpG island found in normal prostate tissues and corresponding cancer samples is shown in Fig. 7. In the normal prostate samples there is a marked transition from extensively methylated DNA upstream of the GSTP1 promoter region to the unmethylated domain of the CpG island and this coincides with the position of an (ATAAA)19–24 repeat. The boundary between the methylated and unmethylated domains in most other tissues examined also correlates well with the position of the ATAAA repeat sequence, even though in brain and smooth muscle the 5'9-boundary position is less marked.

The boundary separating the methylated and unmethylated DNA domains at the 3'9-end of the GSTP1 CpG island is less clearly defined. There is considerable heterogeneity of methylation through a 500-base pair region spanning the end of the CpG island in intron 2 and extending into exons 3 and 4. In prostate cancer the distinct exclusion of methylation from the CpG island is lost, and the CpG island becomes methylated in both GSTP1 alleles.

The nature of the sequences in the boundary region may give a clue as to the mechanism that normally protects the GSTP1 CpG island from methylation. The ATAAA repeat sequence that flanks the 5'9-CpG island is present in about 20 copies and may itself act as a barrier to the methylation of the GSTP1 island in normal cells or could just be fortuitously located 5'9 or 3'9 to the "real" barrier sequence. The sequence immediately adjacent and extending 5' from the ATAAA repeat is a member of the Alu family of interspersed repeated DNA sequences, most closely matching subfamily Sx. The 3'9-end of the Alu sequence corresponds exactly with the start of the ATAAA repeat. Indeed the ATAAA repeat may be an expansion of the residual poly(A) tail of the Alu element. It has been noted previously that unmethylated CpG islands are often flanked by methylated Alu sequences (21). In particular Graff et al. (21)
showed that Alu sequences upstream of the E-cadherin and VHL genes were methylated in normal tissues, whereas adjacent CpG island sequences were not; however, the precise junction of the methylated and unmethylated domains was not determined in this study.

CpG islands are not always flanked by Alu sequences, however, as noted for the 3′-end of the GSTP1 island. Indeed the 3′-boundary of the GSTP1 CpG island, which is diffuse in nature, does not correspond to any identifiable sequence motif. However the start of methylation does correlate with transition from high CpG density of the island to more sparsely spaced CpG sites in the body of the gene. Moreover some of the first CpG site that was found to be methylated (CpG site 31, 1546 base) is located at the junction between intron 1 and exon 2. This location is similar to another boundary of methylation we have previously identified at the 3′-end of the HIC1 CpG island, which also occurs at an exon/intron boundary (22). As for the GSTP1 CpG island the HIC1 intron sequences are unmethylated, whereas adjacent exon sequences show substantial, but not complete methylation.

As discussed above, the boundaries of CpG islands do not appear to harbor common sequence elements at their flanking ends. Indeed the ATAAA repeat sequence boundary found directly 5′ to the GSTP1 CpG island appears to be unique as we have not detected such a structure in other CpG island genes analyzed. However a similar pentanucleotide repeat element has been identified downstream of an Alu sequence in the 3′-untranslated region of a zinc finger cDNA sequence (23). The ATAAA repeat could be fortuitously located 5′ or 3′ to a real barrier sequence, which protects CpG islands from methylation. Alternatively the ATAAA repeat may itself act as a distinct barrier to CpG island methylation of the GSTP1 gene in normal tissues but is in effect in prostate cancer cells. The repeat sequence may provide binding sites for specific protein factors, in particular members of the high mobility group-I(Y) family of mammalian nonhistone proteins. These have been demonstrated to bind specifically to the minor groove of A-T-rich sequences and to function as gene transcriptional regulatory proteins (24). Moreover elevated high mobility group-I(Y) gene expression has been associated with progressive neoplastic transformation (25). Tamimi et al. (26) have shown that in prostate cancer high expression of high mobility group-I(Y) was observed in prostate tumors with higher Gleason grades. It will therefore be of interest to determine if the increased level of high mobility group-I(Y) expression plays a role in the initiation of hypermethylation of the GSTP1 CpG island in prostate cancer.

Whether hypermethylation in the tumor cell is initiated from the ends of the CpG island or at “hypermethylation centers” within the CpG island is unclear. From studies of transfected DNA, Graff et al. (21) have suggested that methylation may progressively encroach from methylated Alu sequence regions flanking CpG islands. Others have suggested that methylation is initiated at “centers” within the islands and progressively spreads (27). The GSTP1 gene from all the separate prostate cancer tissues was found to be extensively methylated throughout the entire CpG island. However, the DNA samples studied were isolated from cells many generations after the initiation of the methylation process. It is therefore difficult to infer how the process may have begun. In the case of the PC3 (Fig. 5) and DU145 (data not shown) prostate cancer cell lines, methylation is found in the core CpG-rich promoter region but does not extend through the 5′-flanking region. This pattern of methylation within the island is also seen in DNA from liver, where GST-π expression has been associated with progressive neoplasia (9, 10). Interestingly it has been noted that high levels of GST-π protein in normal prostate epithelium is limited to the basal cells and that many differentiated secretory epithelial cells do not stain with anti-GST-π antibodies (12). Thus methylation of the GSTP1 gene in prostate cancer cells may be preceded by a loss of expression of the gene in normal epithelium. These observations indicate that early GSTP1 gene inactivation in prostate cancer cells may be a mechanism that predisposes the CpG island promoter region to the de novo methylation pathway resulting in the spread of methylation throughout the island in both alleles.

Acknowledgments—We thank Dr. P Katelaris, S. Danieleitto, and A. Lochhead for support in obtaining prostate tissue samples. DNA from normal tissue samples were kindly provided by Christoff Grunau, Yena.

REFERENCES

1. Schmutz, C., and Jones, P. A. (1998) Biol. Chem. Hoppe-Seyler 379, 377–388
2. Gardiner-Garden, M., and Frommer, M. (1987) J. Mol. Biol. 196, 261–282
3. Antequera, F., and Bird, A. (1995) Proc. Natl. Acad. Sci. U. S. A. 90, 11995–11999
4. Patel, S. A., Graunke, D. M., and Pieper, R. O. (1997) Mol. Cell. Biol. 17, 5813–5822
5. Bird, A. P. (1993) Cold Spring Harbor Symp. Quant. Biol. 58, 281–285
6. Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M., and Issa, J. P. (1998) Advances in Cancer Research 72, 141–196
7. Razaq, A. (1998) EMBO J. 17, 4905–4908
8. Cross, S. H., and Bird, A. P. (1995) Curr. Opin. Genet. Dev. 5, 309–314
9. Brandeis, M., Frank, D., Keshtet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razaq, A., and Cedar, H. (1994) Nature 371, 435–438
10. Macleod, D., Charlton, J., Mullins, J., and Bird, A. P. (1994) Genes Dev. 8, 2283–2292
11. Terrier, P., Townsend, A. J., Cointre, J. M., Triche, T. J., and Cowan, K. H. (1990) Am. J. Pathol. 137, 845–853
12. Cookson, M. S., Reuter, V. E., Linkov, I., and Fair, W. R. (1997) J. Urol. 157, 673–676
13. Lee, W.-H., Morton, R. A., Epstein, J. I., Brooks, J. D., Campbell, P. A., Bova, G. S., Haich, W.-S., Isaacs, W. B., and Nelson, W. G. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 11733–11737
14. Millar, D. S., Ow, K.K., Paul, C. L., Russell, P. J., Molloy, P. L., and Clark S. J. (1999) Oncogene 18, 1313–1324
15. Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994) Nucleic Acids Res. 22, 2989–2997
16. Clark, S. J., and Frommer, M. (1995) in DNA and Nucleoprotein Structure In Vivo (Saluz, H. P., and Wiebauer, K., eds) pp. 123–135, R. G. Landes Co., Austin, Texas
17. Mathew, J., Cattan, A. R., Hall, A. G., Hines, J. E., Nelson, R., Eastham, E., and Burt, A. D. (1992) J. Clin. Pathol. 45, 679–683
18. Mehta, R., Davis, H. G., Laver, G. W., Rowsell, P. R., and Bondy, G. S. (1994) Cancer Lett. 84, 163–172
19. Stirzaker, C., Millar, D. S., Paul, C. L., Warnecke, P. M., Harrison, J., Vincent, P. C., Frommer, M., and Clark, S. J. (1997) Cancer Res. 57, 2229–2237
20. Melki, J. M., Vincent, P. C., and Clark, S. J. (1999) Cancer Res. 59, 3730–3740
21. Graff, J. R., Herman, J. G., Myshaken, S., Baylin, S. B., and Vertino, P. M. (1997) J. Biol. Chem. 272, 22322–22329
22. Melki, J. M., Vincent, P. C., and Clark, S. J. (1999) Leukemia (Baltimore) 13, 877–883
23. Chen, H., Kalaitisadaki, M., Warren, A., Avamopoulos, D., and Antonarakis, S. (1992) Genomics 15, 621–625
24. Friedmann, M., Holth, L. T., Zoghbi, H. Y., and Reeves, R. (1993) Nucleic Acids Res. 21, 4259–4267
25. Ram, T. G., Reeves, R., and Hosick, H. L. (1993) Cancer Res. 53, 2655–2660
26. Tamimi, Y., van der Puij, H. G., Denyse, M. M., Unbes, R., Karthaus, H. F. M., Debruyne, F. M., and Schalken, J. A. (1995) Cancer Res. 55, 5512–5516
27. Toth, M., Lichtenberg, U., and Doerfler, W. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 3728–3732