High-Resolution Methylation Analysis of the Human Hypoxanthine Phosphoribosyltransferase Gene 5' Region on the Active and Inactive X Chromosomes: Correlation with Binding Sites for Transcription Factors

IAN K. HORNSTRA^{1†} AND THOMAS P. YANG^{1,2,3*}

Department of Biochemistry and Molecular Biology,¹ Center for Mammalian Genetics,² and Division of Pediatric Genetics,³ University of Florida College of Medicine, Gainesville, Florida 32610

Received 26 July 1993/Returned for modification 16 September 1993/Accepted 14 November 1993

DNA methylation within GC-rich promoters of constitutively expressed X-linked genes is correlated with transcriptional silencing on the inactive X chromosome in female mammals. For most X-linked genes, X chromosome inactivation results in transcriptionally active and inactive alleles occupying each female nucleus. To examine mechanisms responsible for maintaining this unique system of differential gene expression, we have analyzed the methylation of individual cytosine residues in the 5' CpG island of the human hypoxanthine phosphoribosyltransferase (HPRT) gene on the active and inactive X chromosomes. Methylation analysis of 142 CpG dinucleotides by genomic sequencing was carried out on purified DNA using the cytosine-specific Maxam and Gilbert DNA sequencing reaction in conjunction with ligation-mediated PCR. These studies demonstrate the 5' CpG islands of active and 5-azacytidine-reactivated alleles are essentially unmethylated while the inactive allele is hypermethylated. The inactive allele is completely methylated at nearly all CpG dinucleotides except in a 68-bp region containing four adjacent GC boxes where most CpG dinucleotides are either unmethylated or partially methylated. Curiously, these GC boxes exhibit *in vivo* footprints only on the active X chromosome, not on the inactive X. The methylation pattern of the inactive HPRT gene is strikingly different from that reported for the inactive X-linked human phosphoglycerate kinase gene which exhibits methylation at all CpG sites in the 5' CpG island. These results suggest that the position of methylated CpG dinucleotides, the density of methylated CpGs, the length of methylated regions, and/or chromatin structure associated with methylated DNA may have a role in repressing the activity of housekeeping promoters on the inactive X chromosome. The pattern of DNA methylation on the inactive human HPRT gene may also provide insight into the process of inactivating the gene early in female embryogenesis.

During early mammalian female embryogenesis, one of the two transcriptionally active X chromosomes is randomly inactivated in each cell of the embryo. The inactivation of one X chromosome in each female somatic cell creates a unique system of differential gene expression where a transcriptionally active X chromosome and a transcriptionally inactive X chromosome occupy the same nucleus. The inactivation of genes on one of the two X chromosomes in females compensates for the dosage imbalance of X-linked genes between males and females. The molecular mechanisms that initiate inactivation, propagate the inactivation signal, and maintain this novel system of differential gene expression through subsequent cell divisions are unknown. DNA methylation (21, 27, 28), chromatin structure (22, 38, 41), DNA-protein interactions (13, 31), and DNA replication (12, 13) have all been proposed to have roles in this process.

DNA methylation has been widely implicated in the regulation of gene expression in mammalian cells (3, 42). In many systems of differential gene expression, hypermethylation of certain sites within or flanking genes, particularly in regulatory regions (2, 25), has been correlated with transcriptional silencing (2, 42). DNA methylation in mammals occurs at the cytosine residue of CpG dinucleotides to

produce 5-methylcytosine (26). CpG dinucleotides are generally underrepresented in mammalian genomes but occur at high frequency within CpG islands. These regions in mammalian DNA carry a high G+C content and are often associated with genes, a feature that has been utilized to identify genes by positional cloning (35, 43, 44, 51). CpG islands are often located in the 5' region of constitutively expressed housekeeping genes and are frequently unmethylated in mammalian DNA (2, 26). However, CpG islands associated with the 5' region of housekeeping genes on the inactive X chromosome are characteristically hypermethylated (28, 53–55).

Numerous studies have examined the role of DNA methylation in the process of X chromosome inactivation. Using a variety of experimental approaches, these studies have investigated a correlation between DNA methylation and maintenance of the transcriptionally silent state of genes on the inactive X chromosome (13, 14). These experimental approaches include methylation analysis by methyl-sensitive restriction enzymes in conjunction with Southern blotting (28, 41, 48, 53, 55), DNA-mediated transformation studies using DNA from the active or inactive X chromosomes (27, 49), and analysis of the reactivation of genes from the inactive X chromosome using the DNA-demethylating agent 5-azacytidine (5-azaC) (16, 34, 48, 50). All support the view that the 5' CpG island of housekeeping genes on the inactive X chromosome is hypermethylated in comparison to its corresponding alleles on the active X chromosome. How-

* Corresponding author. Phone: (904) 392-6472. Fax: (904) 392-2953.

† Present address: Division of Dermatology, Washington University School of Medicine, St. Louis, MO 63110.

ever, these studies have not established a consistent correlation between specific sites or levels of DNA methylation in the 5' CpG island and transcriptional repression on the inactive X chromosome (21, 53, 55). Furthermore, a strong correlation between DNA methylation and transcriptional silencing on the inactive X chromosome has not been convincingly established outside of 5' CpG islands (53, 55) nor in X-linked tissue-specific promoters (7). The role of DNA methylation in the process of X inactivation appears to be that of stabilizing the transcriptionally inactive state of CpG-rich promoters following the primary inactivation event (29, 36, 46).

Despite the strong correlation between DNA methylation and silencing of housekeeping genes on the inactive X chromosome, the mechanism by which DNA methylation may repress gene expression on the X chromosome is unclear. Methylation within *cis*-acting regulatory elements may interfere with the binding of *trans*-activating factors to their target sites on DNA (24, 52), but the binding of certain transcription factors such as Sp1 and CTF is unaffected by methylation of their binding sites (1, 17, 18). Methylated DNA may also be a target for DNA-binding proteins that preferentially interact with methylated DNA, thereby repressing transcription of a methylated promoter (32, 33). Alternatively, DNA methylation may suppress transcription by affecting chromatin structure (5, 23). Recent evidence suggests that methylation within the preinitiation domain of the promoter exhibits the strongest correlation with repression of promoter activity (25). Thus, specific sites or regions within the promoter may be crucial for repressing transcription of genes on the inactive X chromosome by DNA methylation.

Recently, Pfeifer et al. (40, 41) have examined the methylation of individual cytosine residues in the 5' CpG island of the X-linked human phosphoglycerate kinase (PGK-1) gene. They have employed the high-resolution technique of ligation-mediated PCR (LMPCR) genomic sequencing to determine the methylation state of each and every CpG dinucleotide on the active and inactive X chromosome. This method overcomes the significant limitations of methylation analysis using methylation-sensitive restriction enzymes in conjunction with Southern blot analysis. Methylation-sensitive restriction enzymes assay only a fraction of all CpG dinucleotides and often do not permit precise mapping of methylated and unmethylated restriction sites in regions with a high density of closely spaced restriction sites (such as CpG islands), particularly if the region is partially methylated or unmethylated. Genomic sequencing permits direct examination of the methylation state of all cytosines regardless of methylation status and allows the determination of a comprehensive high-resolution methylation pattern within a specific region of genomic DNA. Pfeifer et al. (40, 41) found the active PGK-1 allele was completely unmethylated at 120 CpG sites on the active X chromosome but was essentially completely methylated (118 of 120 CpG sites) on the inactive X chromosome.

Hypoxanthine phosphoribosyltransferase (HPRT; EC 2.4.2.8) catalyzes the conversion of hypoxanthine and guanine to IMP and GMP, respectively, in the purine salvage pathway. The HPRT gene is constitutively expressed in all cells and tissues throughout development with elevated expression in the central nervous system, particularly the basal ganglia (47). The HPRT gene is X linked and transcriptionally silenced on the inactive X chromosome. We have previously employed *in vivo* footprinting to identify the positions of multiple sequence-specific DNA-protein inter-

actions specific to the 5' CpG island of the active HPRT allele; no *in vivo* footprints were detected on the inactive allele (19).

Previous methylation analysis of the human HPRT gene using methylation-sensitive restriction enzymes suggests that, unlike the PGK-1 gene, the 5' CpG island on the inactive X chromosome is not completely methylated (53, 55). Therefore, we have analyzed the human HPRT gene 5' CpG island by LMPCR genomic sequencing to determine the complete methylation pattern on the active and inactive X chromosomes. This high-resolution map of individual methylated and unmethylated cytosines was then correlated with transcriptional activity of the gene and the pattern of binding sites for transcription factors that interact with the promoter region *in vivo* (19). We find a nearly complete absence of DNA methylation on active and 5-azaC-reactivated HPRT alleles. The inactive allele is nearly completely methylated at all CpG dinucleotides, except in a region containing four adjacent GC boxes which has been shown by *in vivo* footprinting to be bound by sequence-specific DNA binding proteins only on the active allele. CpG dinucleotides in this region are either partially methylated or unmethylated in two independent cell lines carrying an inactive X chromosome. These data provide insight into molecular processes that may be involved in X chromosome inactivation.

MATERIALS AND METHODS

Cell lines. DNA samples were prepared from cultures of cell lines previously described (19). Briefly, GM00468 is a normal diploid human male fibroblast cell line containing an active X chromosome. Cell line 4.12 (generously provided by David Ledbetter) is a hamster-human somatic cell hybrid containing only the active human X chromosome in the HPRT-deficient hamster cell line RJK88; RJK88 carries a deletion of the endogenous hamster HPRT gene (10). Cell line 8121 is a hamster-human somatic cell hybrid containing an inactive human X chromosome in a RJK88 hamster cell background (also provided by David Ledbetter). Cell line 8121R9a is a 5-azaC reactivant of 8121 grown from a single hypoxanthine-aminopterin-thymidine-resistant colony expressing the 5-azaC-reactivated human HPRT gene. In some experiments, a second 5-azaC reactivant was studied; cell line M22 is a 5-azaC-treated HPRT reactivant of a mouse-human somatic cell hybrid containing an inactive human X chromosome in a murine A9 cell background (generously provided by Barbara Migeon). An additional cell line, X8-6T2, is a hamster-human somatic hybrid cell line containing an inactive human X chromosome (8, 9, 15) (generously provided by Stanley Gartler) and grown in Dulbecco modified Eagle medium with 10% fetal bovine serum and 1% penicillin-streptomycin.

DNA preparation and base-specific modification. Genomic DNA from each cell line was isolated as previously described (19). LMPCR genomic sequencing was performed as described by Hornstra and Yang (20). This is a modification of the original genomic sequencing method described by Church and Gilbert (30). Briefly, purified genomic DNA (50 µg) was digested with *Eco*RI to decrease viscosity, phenol:chloroform (50:50) extracted, and ethanol precipitated. The digested DNA was resuspended in 5 µl of water and 15 µl of 5 M NaCl and then subjected to the standard Maxam and Gilbert cytosine-specific modification reaction with hydrazine (30). Hydrazine modification of 50 µg of genomic DNA for 16 min at room temperature was found to be optimal. After cleavage of the DNA at hydrazine-modified cytosines

by piperidine treatment (30), 1/10 volume of 3 M sodium acetate (pH 7) was added, and the DNA was precipitated with 2 volumes of ethanol and collected by centrifugation at $14,000 \times g$ for 30 min. After the supernatant was decanted, the pellet was washed twice with 80% ethanol and dried overnight in a vacuum concentrator. The chemically cleaved genomic DNA was resuspended in $1 \times$ TE (10 mM Tris [pH 8], 1 mM EDTA) at approximately $1 \mu\text{g}/\mu\text{l}$.

For controls, 10 μg of plasmid DNA, which contains a 1.8-kb fragment of human HPRT 5' region, was linearized with *Eco*RI and subjected to each of the four standard Maxam and Gilbert sequencing reactions (G, A+G, T+C, C) (30). After vacuum drying, the plasmid samples were diluted to a final concentration that would produce signals in the final autoradiogram equal in intensity to that of a single-copy mammalian gene after LMPCR of genomic DNA.

LMPCR. LMPCR was carried out as described by Hornstra and Yang (19, 20) with a modification of the Garrity and Wold procedure (11) employing Vent DNA polymerase (New England Biolabs). For the LMPCR, six primer sets previously described for *in vivo* footprinting of the human HPRT gene were used (19), as well as two new primer sets, I and J: primer I1 (5'-HO-TTGCTGCGCTCCGCTC-OH-3') and primer I2 (5'-HO-CGGCTTCCTCCTCTGAGCA GTCA-OH-3'); primer J1 (5'-HO-CGCCATTTCCACCTT CTCTT-OH-3') and primer J2 (5'-HO-TTCCACACGCA GTCCTCTTTTCCCA-OH-3').

For primer extension (first-strand synthesis) with Vent DNA polymerase, 1 to 5 μg of chemically cleaved (hydrazine- and piperidine-treated) genomic DNA (or the equivalent copy number of treated plasmid DNA), 0.6 pmol of primer 1, 3 μl of $5 \times$ Vent buffer ($5 \times$ Vent buffer is 200 mM NaCl plus 50 mM Tris-HCl, pH 8.9) were mixed, and water was added to bring the total volume to 15 μl . This mixture was incubated at 98°C for 10 min to denature the DNA, followed by annealing of the primer at 45°C for 30 min. The samples were cooled on ice, and 15 μl of a freshly prepared solution was added to each tube to yield a solution with a final concentration of 40 mM NaCl, 10 mM Tris-HCl (pH 8.9), 5 mM MgSO_4 , 0.25 mM 7-deaza-dGTP-deoxynucleotide triphosphate (dNTP) mix (0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dTTP, 0.1875 mM 7-deaza-dGTP, 0.0625 mM dGTP), and 2 U of Vent DNA polymerase. The first strand synthesis (primer extension) was incubated at 53°C for 1 min, 55°C for 1 min, 57°C for 1 min, 60°C for 1 min, 64°C for 1 min, 68°C for 1 min, 72°C for 3 min, and 76°C for 3 min, and then the tubes were placed on ice. Twenty microliters of dilution solution (11) was added, followed by 25 μl of the ligation solution described by Garrity and Wold (11). The samples were incubated at 17°C overnight for ligation. After the ligation, 40 μl of 7.5 M ammonium acetate and 1 μl of a 10-mg/ml tRNA solution was added to each tube and ethanol precipitated by the addition of 2 volumes of ethanol. The DNA was collected by centrifugation, the supernatant was decanted, the pellet was washed with 80% ethanol, and the pellet was dried under vacuum. The dried pellet was then redissolved in 20 μl of water. For PCR amplification, 80 μl of a PCR solution was added so the final concentration in the 100- μl PCR reaction was $1 \times$ Vent buffer, 3 mM MgSO_4 , 0.25 mM 7-deaza-dGTP-dNTP mix, 25 pmol of primer 2, 20 pmol of the 25-mer linker primer, and 3 U of Vent DNA polymerase. Eighty microliters of mineral oil was added to each tube, and the samples were placed in a temperature cycler (Coy II) for the PCR. The samples were initially denatured at 95°C for 3 min, and then the samples were repetitively denatured at 95°C for 1 min, annealed at 66°C for 2 min, and

extended at 76°C for 3 min; the samples were cycled in this manner 20 times. Additionally, with each five cycles, the extension time was increased 30 s. After 20 cycles, the tubes were incubated at 76°C and 5 μl of a booster solution (containing $1 \times$ Vent buffer, 3 mM MgSO_4 , 5 mM dATP, 5 mM dCTP, 5 mM dGTP, 5 mM dTTP, and 1 U of Vent DNA polymerase) was added to each sample. The samples were incubated at 76°C for 10 min to allow Vent DNA polymerase to complete the formation of blunt ends. The samples were placed on ice, and 3 μl of 0.5 M EDTA was added. Subsequent gel electrophoresis and electroblotting were carried out as previously described, using a 5% Long Ranger gel (AT Biochem) substituted for the standard polyacrylamide DNA sequencing gel (19). To visualize the final DNA sequencing ladder, single-stranded hybridization probes were synthesized from M13 clones containing the human HPRT promoter region cloned in either orientation. Probe synthesis, hybridization, washing, and autoradiography were performed as previously described (19, 20).

RESULTS

The methylation state of every detectable cytosine in the 5' CpG island of the human HPRT gene was directly examined by the method of genomic sequencing. The 730-bp region spanning positions -530 to +202 (relative to the translation initiation codon) on both the active and inactive X chromosomes was subjected to genomic sequencing analysis using the LMPCR technique (11). This region contains the 5' flanking region, as well as the first exon and the 5' portion of the first intron, and includes most of the 5' CpG island.

The analysis was performed on six different cell lines to examine the methylation state of each cytosine nucleotide on either the active or inactive HPRT allele. The active allele was analyzed in a rodent-human somatic cell hybrid containing an active human X chromosome (cell line 4.12), in a diploid human male fibroblast cell line (GM00468), and in somatic cell hybrids containing a 5-azaC-reactivated HPRT gene on the inactive X chromosome (cell lines 8121R9a and M22). The inactive allele was analyzed in two independently-derived somatic cell hybrids containing an inactive human X chromosome (cell lines 8121 and X8-6T2).

The application of LMPCR to the original genomic sequencing method of Church and Gilbert (6) overcomes the technical limitations previously encountered when analyzing single-copy genes within the DNA of highly complex genomes (37, 40). LMPCR amplifies each DNA fragment in sequencing ladders from a specific region of interest within genomic DNA after chemical cleavage of the DNA by each of the base-specific Maxam and Gilbert reactions (30). Methylation analysis by genomic sequencing is based upon the resistance of 5-methylcytosine residues to hydrazine modification in the cytosine-specific Maxam and Gilbert reaction. As a result, 5-methylcytosine residues within genomic DNA are indicated by bands missing from the cytosine sequencing ladder but present in the ladder from an unmethylated sample.

For our analysis, an individual cytosine residue was considered to be methylated if the intensity of the band in the sequencing ladder was visually estimated to be less than 25% the intensity of the same band in an unmethylated sample (active X genomic DNA or plasmid DNA containing the human HPRT gene 5' region). Partially methylated cytosines were those that exhibited approximately 25 to 80% of the unmethylated band intensity, and unmethylated cytosines

FIG. 1. Locations of primers used in LMPCR genomic sequencing analysis of the human HPRT 5' region. The numbered line represents the human HPRT 5' region with positions numbered relative to the translation initiation codon. The large rectangle represents the first exon, with the cross-hatched portion signifying the region of multiple transcription start sites. The small solid rectangles above and below the numbered line indicate positions of the PCR primer sets used for the LMPCR genomic sequencing. Primer sets N, A, M, and I are complementary to the lower strand sequence and analyze the lower strand; primers E, C, R, and J are complementary to the upper strand sequence and analyze the upper strand. Lines with arrowheads indicate the region resolved by each primer set.

were those deemed to possess greater than 80% of the control band intensity by visual inspection. Partially methylated sites occur at specific CpG dinucleotides that are methylated in some cells and unmethylated in others within the same sample.

Figure 1 shows the relative positions of the oligonucleotide primer sets and the region covered by each primer set for LMPCR genomic sequencing of the human HPRT gene 5' region. The region between positions -530 to +202 was analyzed for cytosine methylation on both strands.

Primer sets N, A, M, and I were used to analyze the lower strand of the HPRT 5' region, and primer sets J, E, C, and R were used to examine methylation of the upper strand. Representative cytosine-specific genomic sequencing ladders generated by primer sets A, M, E, and C are shown in Fig. 2, 3, 4, and 5, respectively. Data from genomic sequencing ladders generated by primer sets N, I, J, and R are not shown, but results from these primer sets are included in the methylation summary diagrams in Fig. 6 and 7.

Analysis of the lower strand. Results of LMPCR genomic sequencing of the lower strand from position -411 to -253 by using primer set A is shown in Fig. 2. Within this region, all cytosine residues in CpG dinucleotides are unmethylated in cell lines containing an active or 5-azaC-reactivated HPRT gene (Fig. 2, lanes 1, 2, 5, and 6). This was determined by comparing the cytosine band intensities of these samples with those in a similar cytosine-specific LMPCR genomic sequencing ladder from purified plasmid DNA containing the human HPRT 5' region (plasmid ladder not shown); bacterial plasmid DNA is not methylated at CpG dinucleotides. On the active allele, the relative intensity of all cytosine bands from CpG dinucleotides was the same for the plasmid DNA and the genomic DNA samples containing an active X chromosome.

Analysis of the two somatic cell hybrids containing an inactive human X chromosome (Fig. 2, lanes 3 and 4) shows hypermethylated cytosines at all CpG dinucleotides in the region covered by this primer set. For example, the cytosine at position -372 displays strong bands in the two samples containing active X chromosomes (lanes 1 and 2), indicating a lack of significant methylation, and exhibits significantly

less intense bands in the two samples containing an inactive X chromosome (lanes 3 and 4). In cell line 8121 (lane 3), the band intensity is significantly reduced (compared with that of the unmethylated samples containing the active X in lanes 1 and 2) but is still readily detectable, indicating a partially methylated cytosine at this position in this cell line. However, cell line X8 shows no band detectable above the faint background ladder (lane 4), indicating that the cytosine at position -372 in this sample is completely methylated. In both cell lines where the inactive human HPRT gene has been reactivated by 5-azaC treatment (8121R9a and M22), the relative band intensity indicative of an unmethylated cytosine is restored (lanes 5 and 6).

Results of LMPCR genomic sequencing of the lower strand from position -233 to -53 with primer set M is shown in Fig. 3. Again, all CpG dinucleotides in this region are unmethylated in both cell lines containing an active HPRT allele (Fig. 3, lanes 1 and 2), as well as in the 8121R9a 5-azaC reactivant (lane 5) and in 5-azaC reactivant M22 (data not shown). In the two samples containing an inactive human X chromosome, all CpG dinucleotides are completely methylated in the region between positions -53 to -139 (Fig. 3, lanes 3 and 4). However, immediately upstream of this region in these samples, between positions -164 and -233, all CpGs are either completely unmethylated or partially methylated on the inactive X chromosome (lanes 3 and 4). This cluster of hypomethylated sites on the inactive X chromosome coincides with the location of four GC boxes (marked I, II, III, and IV in Fig. 3) which exhibit *in vivo* footprints on the active HPRT allele (19). Curiously, no *in vivo* footprints have been detected in this region on the inactive allele. In cell line 8121, the region containing the four GC boxes (positions -164 to -219) on the lower strand consists entirely of unmethylated sites (Fig. 3, lane 3). But further downstream of position -164 in this cell line, nearly all of the CpG dinucleotides return to the completely methylated state. Similarly, in cell line X8-6T2, the same GC box region on the lower strand contains an interspersed pattern of unmethylated, partially methylated, and completely methylated sites (Fig. 3, lane 4). Again, further downstream of

FIG. 2. Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set A. The autoradiogram shows the cytosine-specific sequencing ladder from position -411 to position -253 (relative to the translation initiation codon). The position relative to the translation initiation codon is shown to the right of the sequencing ladder. The horizontal bars to the left of the sequencing ladder indicate the position of cytosines in CpG dinucleotides. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, normal diploid male fibroblasts (cell line GM00468); lane 2, hamster-human somatic cell hybrid cells containing the active human X chromosome (cell line 4.12); lane 3, hamster-human somatic cell hybrid cell containing the inactive human X chromosome (cell line 8121); lane 4, hamster-human somatic cell hybrid cells containing the inactive human X

position -164 in this cell line, nearly all of the CpG dinucleotides are completely methylated.

Results from analysis of the lower strand from position -12 to position +128 using primer set I (data not shown) indicate both cell lines carrying an active X chromosome (GM00468 and 4.12) as well as both cell lines with a 5-azaC reactivated HPRT gene (8121R9a and M22) are unmethylated at all CpG dinucleotides. In both cell lines carrying an inactive X chromosome (8121 and X8), methylation of CpGs is nearly complete in this region.

Methylation analysis of the 4 CpG dinucleotides between positions -411 to -446 with primer set N yields one unusual methylation pattern (data not shown). Though all four CpG dinucleotides in the male fibroblast cell line (GM00468) are completely unmethylated, 2 of the 4 sites in the hybrid cell line carrying an active X chromosome (4.12) are partially methylated (at positions -425 and -427), and the remaining two sites (positions -411 and -446) are unmethylated. Both cell lines carrying a 5-azaC-reactivated HPRT gene (8121R9a and M22) show no methylation at any of these sites. On the inactive X chromosome in hybrid 8121, these four CpG dinucleotides are either partially or completely methylated; hybrid cell line X8 was not examined in this region.

Analysis of the upper strand. Genomic sequencing of the upper strand shows the same methylation patterns found on the lower strand. Results of LMPCR genomic sequencing from positions -10 to -138 using primer set E are shown in Fig. 4. On the active and 5-azaC-reactivated alleles (lanes 1, 2, and 5), all cytosines in CpG dinucleotides are unmethylated. In both somatic cell hybrids containing an inactive X chromosome, all CpGs are completely methylated (lanes 3 and 4).

The region spanning positions -145 to -289 was examined on the upper strand by using primer set C (Fig. 5). This region contains the four GC boxes (-164 to -233) which are shown in Fig. 5 as I, II, III, and IV. Again, the region is completely unmethylated on the active and 5-azaC-reactivated alleles (lanes 1, 2, 5, and 6). In both cell lines carrying an inactive X chromosome, most CpG dinucleotides in the GC box region are unmethylated (lane 3) or interspersed with methylated, partially methylated, and unmethylated sites (lane 4). Upstream of the GC boxes, the pattern of hypermethylation typically found on the inactive X chromosome is restored.

Analysis of the upper strand from position -383 to -447 was performed by using primer set R (data not shown). All eight CpG dinucleotides in the normal male cell line (GM00468) are unmethylated in this region. However, in the somatic cell hybrid carrying an active X chromosome (4.12), two of the eight CpGs are partially methylated at positions -426 and -428, while the remaining six sites are unmethylated. The two partially methylated sites in this cell line correlate with the position of the partially methylated sites seen on the lower strand in this cell line (see above) by using primer set N. Analysis of the inactive HPRT allele in cell lines 8121 and X8 carrying the inactive human X chromosome shows either total or partial methylation at every CpG

chromosome (cell line X8-6T2); lane 5, hamster-human somatic cell hybrid cells containing a 5-azaC-reactivated human HPRT gene on the inactive human X chromosome (cell line 8121R9a); lane 6, mouse-human somatic cell hybrid cells containing a 5-azaC-reactivated human HPRT gene on the inactive X chromosome (cell line M22).

FIG. 3. Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set M. The autoradiogram shows the cytosine-specific sequencing ladder from -232 to -53. The symbols and designations are identical to those in Fig. 2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; and lane 5, cell line 8121R9a. Methylation data from cell line M22 are not shown. The brackets and roman numerals on the left indicate GC boxes I, II, III, and IV and correlate with the GC boxes labeled in Fig. 6 and 7.

FIG. 4. Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set E. The autoradiogram shows the cytosine-specific sequencing ladder from -10 to -134. The symbols and designations are identical to those in Fig. 2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; and lane 5, cell line 8121R9a.

FIG. 5. Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set C. The autoradiogram shows the cytosine-specific sequencing ladder from -145 to -289. The symbols and designations are identical to those in Fig. 2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; lane 5, cell line 8121R9a; and lane 6, cell line M22. The brackets and roman numerals on the left indicate GC boxes I, II, III, and IV and correlate with the GC boxes labeled in Fig. 6 and 7.

dinucleotide. This region in both cell lines containing a 5-azaC-reactivated HPRT gene is completely unmethylated.

The upper strand from position +202 to +24 was analyzed by using primer set J (data not shown). In this region, the active and 5-azaC-reactivated HPRT alleles again are completely unmethylated at all CpG dinucleotides. The inactive HPRT allele is completely methylated at all CpGs in cell line X8 and methylated at all CpGs in cell line 8121 except at position +186, which is completely unmethylated, and position +194, which is partially methylated.

Summary of methylation analysis. The methylation pattern of the human HPRT gene on the active X chromosome was examined in two different cell lines. In the diploid male fibroblast (GM00468), the HPRT gene is completely unmethylated at 142 of 142 CpG dinucleotides assayed; the methylation state of 10 additional sites could not be determined because of technical limitations of the LMPCR genomic sequencing (where some cytosines are not resolvable or detectable in the DNA sequencing ladder). Somatic cell hybrid 4.12 containing the active X chromosome is unmethylated at 138 of 142 sites and partially methylated at a cluster of 4 CpG dinucleotides at the far 5' end of the region analyzed (at positions -426 and -428 on the upper strand and -425 and -427 on the lower strand).

The inactive HPRT allele was examined in two different somatic cell hybrids containing an inactive human X chromosome. The methylation patterns of the inactive HPRT gene in these two cell lines are summarized in Fig. 6 and 7. In cell line 8121, 107 of 142 CpG dinucleotides are completely methylated, 9 CpGs are partially methylated, and 26 CpGs are unmethylated. Twenty-four of the 26 unmethylated sites in hybrid 8121 are located in the region of the four GC boxes between positions -233 and -164. In hybrid cell line X8-6T2, 122 of 142 CpG dinucleotides are methylated, 12 CpGs are partially methylated, and 9 CpGs are completely unmethylated. All 9 completely unmethylated sites and 6 of the 12 partially methylated sites are located in the region of the GC boxes in cell line X8-6T2. Thus, in two independent cell lines carrying an inactive human X chromosome, the region containing the four GC boxes is hypomethylated (with unmethylated and partially methylated sites) relative to the surrounding region of the 5' CpG island.

Reactivation of the HPRT gene on the inactive X chromosome by treatment of cells with 5-azaC demethylates all CpG dinucleotides; cell lines 8121R9a and M22 were completely unmethylated at all 142 of 142 CpG dinucleotides in the 5' region. Thus, 5-azaC reactivation of the human HPRT gene on the inactive X chromosome restored the methylation pattern to a pattern indistinguishable from that of the active HPRT allele.

DISCUSSION

Methylation analysis of the human HPRT gene by genomic sequencing provides high-resolution data that further refines previous methylation analysis by methyl-sensitive restriction enzymes (53, 55). Our genomic sequencing studies have focused exclusively on the methylation status of the 5' CpG island and permit an examination of the methylation state of every cytosine nucleotide in the region on active and inactive X chromosomes. This method yields precise and definitive information on the methylation patterns of the active and inactive alleles not available by studies with methylation-sensitive restriction enzymes.

Overall, results from our methylation analysis by genomic sequencing are consistent with previous methylation analy-

FIG. 6. Summary of the methylation pattern of cytosines from the human HPRT 5' region on the inactive X chromosome in hybrid cell line 8121. The sequence of the human HPRT 5' region is shown. Numbers on the right side of the sequence indicate the position relative to the translation initiation codon marked as +1. The thick solid line underlines the coding region of exon 1. The thin dashed line indicates the region of multiple transcription initiation sites. The GC boxes (which exhibit in vivo footprints on the active HPRT allele) are indicated by a thin solid lined and marked by roman numerals I, II, III, and IV. Guanine residues that are footprinted in vivo by dimethyl sulfate on the active HPRT allele (19) are shown in bold italics. Solid circles denote methylated cytosine residues. Partially filled circles indicate partially methylated cytosine residues. Open circles represent unmethylated cytosine residues. Question marks indicate cytosine residues which could not be resolved in the DNA sequencing ladder or whose methylation status could not be determined.

sis using restriction enzymes in conjunction with Southern blotting (53, 55). These previous studies indicated that active HPRT alleles are extensively hypomethylated at restriction sites in the 5' region relative to inactive alleles. However, in part because of technical limitations of these earlier studies, no consistent pattern of methylation at these sites could be discerned and correlated with silencing of the HPRT gene on the inactive X chromosome, particularly within the 5' CpG island. Our analysis by genomic sequencing demonstrates a near total absence of methylation on the active HPRT 5' CpG island in male fibroblast DNA as well as in a somatic cell hybrid bearing the active human X chromosome and in 5-azaC-reactivated HPRT alleles. The inactive allele in two independent somatic cell hybrids shows a very clear pattern of hypermethylated CpG dinucleotides surrounding a short

FIG. 7. Summary of the methylation pattern of cytosines from the human HPRT 5' region on the inactive X chromosome in hybrid cell line X8-6T2. All symbols are identical to those in Fig. 6.

(48 to 68 bp) tract of variably hypomethylated sites within the CpG island. These data suggest that some of the heterogeneity in the methylation pattern of the 5' region on inactive HPRT alleles found by using restriction enzymes (53, 55) may be due, in part, to this variably hypomethylated region. We have not analyzed the methylation pattern of the 5' CpG island in diploid female cells because of the inability to separate the active and inactive HPRT alleles in these samples.

Correlation of methylation patterns with binding sites for transcription factors. In vivo footprint analysis of the human HPRT gene 5' CpG island on the active and inactive X chromosomes has demonstrated multiple footprints specific to the active HPRT allele; no in vivo footprints were detected on the inactive allele (19). The in vivo footprint pattern on the active allele includes evidence for binding of transcription factors to four adjacent GC boxes (positions -163 to -215), DNA sequences known to interact with the transcription factor Sp1 (4). In addition, the active allele exhibits in vivo footprints at a potential AP-2 binding site (from -265 to -286) and at a position just downstream of the multiple transcription initiation sites that may define the binding site of a new transcription initiation factor (from positions -75 to -91). The positions of these in vivo footprints within the CpG island are indicated on Fig. 6 and 7.

On the active HPRT allele, the 5' CpG island is completely unmethylated at all CpG dinucleotides within the DNA

sequences of all *in vivo* footprints and in the region of the multiple transcription start sites. This absence of methylated cytosines correlates with the binding of transcription factors and transcriptional activity.

The 5' CpG island of inactive HPRT allele, which lacks any evidence for *in vivo* footprints, is extensively methylated. Figures 6 and 7 summarize methylation analysis of the inactive allele in two different somatic hybrid cell lines carrying an inactive X chromosome. Comparison of the methylation pattern on the inactive alleles with the pattern of *in vivo* footprints on the active allele reveals an interesting correlation. The region of the 5' CpG island bearing the four adjacent GC boxes is hypomethylated relative to the surrounding regions of the CpG island on the inactive allele, with hybrid cell line 8121 methylated to a lesser extent in this region than hybrid cell line X8-6T2. In cell line 8121, the GC box region is completely unmethylated at all CpGs, while in cell line X8-6T2, the GC box region is interspersed with unmethylated, partially methylated, and fully methylated CpGs. The molecular basis for this tract of hypomethylated CpGs on the inactive allele is not clear, though some speculation is possible (see below). This methylation pattern on the inactive allele is particularly unusual because the inactive allele is not bound *in vivo* by sequence-specific binding proteins (19), and the binding of the transcription factor Sp1 to GC box sequences has been shown to be unaffected by CpG methylation within the binding sequence (17, 18). Thus, the only hypomethylated region in the 5' CpG island on the inactive allele occurs within unoccupied binding sites for a transcription factor that is not affected by methylation of its DNA target. However, similar studies of the mouse HPRT gene 5' CpG island show no obvious correlation between hypomethylated sites and binding sites for transcription factors (27a).

Pfeifer et al. (39) have also demonstrated that partial demethylation of the 5' CpG island of the inactive PGK-1 gene (which normally is methylated at all CpGs on the inactive X chromosome) by 5-azaC treatment is not sufficient to reactivate expression of the inactive allele, even when only 50% of CpG sites remain methylated. Analysis of three independent 5-azaC-treated isolates which do not express PGK-1 exhibit different patterns of interspersed methylated and unmethylated sites. However, the methylation patterns in these isolates suggest that hypermethylation at or near the transcription start site may be critical for transcriptional repression. These results are consistent with our findings on the methylation of the human HPRT gene on the inactive X chromosome. Furthermore, similar 5-azaC demethylation experiments on the human HPRT gene (i.e., analysis of non-HPRT-expressing isolates) may provide further information on which sites or regions of the human HPRT gene 5' CpG island are critical for repression by methylation. These experiments are currently under way.

One explanation for the hypomethylated GC box region on the inactive HPRT gene may lie in the fact that the region of the four GC boxes has a high incidence of GCG and CGC trinucleotides. DNA methyltransferase may have a bias against methylation of GCG and CGC trinucleotides (41) which would leave these sites hypomethylated in genomic DNA from the inactive HPRT allele. A methylation pattern consistent with this possibility has been noted in the inactive human PGK-1 gene; Pfeifer et al. observed that CGC and GCG trinucleotides are often partially methylated on the inactive PGK-1 allele (41). Examination of the hypomethylated GC box region of the HPRT gene on the inactive X chromosome indicates that most unmethylated or partially

methylated sites in this region occur within CGC or GCG trinucleotides. However, these trinucleotide sequences are also frequently found at fully methylated sites on the inactive alleles, and not all unmethylated or partially methylated sites occur within CGC or GCG trinucleotides.

DNA methylation in the GC box region is unlikely to be directly responsible for modulating the differential binding of Sp1 on the active and inactive alleles because this region is hypomethylated on inactive HPRT alleles and because of the ability of Sp1 to bind methylated binding sites (17, 18). However, it is possible that methylation could directly affect the binding of other transcriptional activators (at other *in vivo* footprinted sites on the active allele) by lowering the affinity of the proteins for their binding site on the inactive allele. For example, the *in vivo* footprinted region involving position -91 (19) is associated with a high density of CpG dinucleotides that are differentially methylated on the active and inactive X chromosomes on both the upper and lower strands; all CpGs in this region are completely unmethylated on the active allele and completely methylated on the inactive allele. It is possible that sequence-specific DNA-binding proteins interacting in this region may be affected by methylation of their binding sites. Methylation of the -91 *in vivo* footprint region may act to repress transcription of the HPRT gene on the inactive allele because proteins associated with this region may be involved in formation of the preinitiation complex (19), and Levine et al. report that methylation in the preinitiation domain is most effective in suppressing promoter activity (25).

Further upstream in the 5' CpG island at the potential AP-2 site (or adenoviral E2aE-CB and E4E2 sites) near position -266, two nearby CpGs are also differentially methylated. The two sites are totally unmethylated on the active allele and either partially methylated or fully methylated on the inactive allele. The effect of methylation at this site and in this region is unknown.

Comparison of the high-resolution methylation patterns of the human HPRT and PGK-1 5' CpG islands. Comparison of the methylation pattern from the human HPRT gene with the pattern obtained by Pfeifer et al. (41) from the X-linked human PGK-1 gene reveals nearly identical patterns on the active alleles. On the active alleles, both genes are unmethylated at CpG dinucleotides; the PGK-1 gene on the active X chromosome is unmethylated at each of 120 CpGs, and the HPRT gene is unmethylated at each of 142 CpGs in male fibroblasts and in 5-azaC-reactivated HPRT genes and unmethylated at 138 of 142 CpGs in a somatic cell hybrid carrying an active X chromosome. Thus, transcriptional activity of these X-linked housekeeping genes correlates with an essentially unmethylated 5' CpG island.

On the inactive allele of the HPRT and PGK-1 genes, the general level of methylation is similar (both are hypermethylated relative to the active alleles), but the pattern of methylation is strikingly different. Comparison of the methylation pattern and *in vivo* footprint pattern of the PGK-1 gene yields no obvious correlation between unmethylated, methylated, or partially methylated sites and the location of binding sites for sequence-specific DNA-binding protein (41). Furthermore, GC box regions in the PGK-1 gene do not show an unusually high incidence of unmethylated or partially methylated sites. However, examination of the human HPRT gene on the inactive X chromosome shows a clear correlation between the GC boxes (which exhibit *in vivo* footprints only on the active allele) and a cluster of unmethylated and partially methylated sites. It should be noted that the same X8-6T2 hybrid cell line was used in genomic

sequencing studies of both the HPRT and PGK-1 genes on the inactive X chromosome. Thus, the difference in methylation patterns between the PGK-1 and HPRT genes on the inactive X chromosome is not simply due to a difference in the cells studied.

Hypermethylation is correlated with the maintenance of transcriptional repression, but as evidenced by the unmethylated and partially methylated sites on the inactive allele of the HPRT gene, complete methylation of the 5' CpG island is not required for silencing all housekeeping genes on the inactive X chromosome. Thus, the specific position of methylated CpG dinucleotides, the overall density of methylation, and/or the length of methylated regions in the 5' CpG island may be critical for maintaining the transcriptionally suppressed state of housekeeping genes on the inactive X chromosome.

Implications for X chromosome inactivation. Hypomethylation of the GC box region on the inactive X chromosome suggests a sequence of events that may occur on the HPRT 5' CpG island early in female embryogenesis at the time of X chromosome inactivation. Transcriptional silencing of the HPRT gene appears to occur prior to de novo methylation of CpG dinucleotides in the 5' CpG island (29, 36, 46). If transcriptional activator proteins bound to the GC box region (most likely Sp1) are not displaced at the time of inactivation of the HPRT gene and prior to de novo methylation of the 5' CpG island, the continued presence of the bound proteins may protect CpG dinucleotides within the binding site from methylation. The delay in displacing transcription factors in the region of the GC boxes in all or some cells during the X inactivation process would allow CpG dinucleotides covered by the binding proteins to escape methylation, resulting in unmethylated and partially methylated CpG sites in the region occupied by transcription factors on the active allele. Once this pattern of methylation on the inactive X chromosome is established early in embryogenesis, this pattern would persist into adult cells via the maintenance DNA methylase and would yield the hypomethylated GC box region seen in the two somatic cell hybrids carrying the inactive X chromosome. Presumably, proteins binding to the GC box region would be released or displaced sometime after DNA methylation, since we observed no footprints in this region of the HPRT gene on the inactive X chromosome in our previous *in vivo* footprinting studies (19).

This scenario would also imply that simultaneous displacement of all transcriptional activators from X-linked genes undergoing inactivation does not necessarily occur at the time of X inactivation, and that displacement of certain key transcription factors may occur first and may be all that is initially required for inactivation of some X-linked genes. In the case of the human HPRT gene, this key factor(s) may be binding to the region surrounding position -91 as seen by previous *in vivo* footprinting studies (19). This region shows complete methylation in both cell lines carrying the inactive X chromosome. Levine et al. (25) have reported that the most effective repression of genes by DNA methylation was observed when methylation occurred in the preinitiation domain of the promoter. The position of the -91 footprint region and the absence of a TATA box in the HPRT gene suggests that this DNA-protein interaction may be involved in formation of the preinitiation complex. Thus, displacement of factors from the -91 region followed by methylation of this region by X chromosome inactivation may be sufficient to inactivate the HPRT gene during female embryogenesis. However, because the human PGK-1 gene shows no

evidence for *in vivo* footprints in the region analogous to the -91 footprint region of the HPRT gene, displacement of Sp1 from the GC box region (prior to DNA methylation) may be crucial for silencing of the PGK-1 gene at the time of X inactivation. This could account for the difference in methylation patterns of the GC boxes in the HPRT and PGK-1 5' CpG islands.

Because there is no obvious and consistent correlation between sites of methylated CpG dinucleotides and binding sites for DNA-binding regulatory proteins, methylation of the 5' CpG island of housekeeping genes may be involved in stabilizing the chromatin structure of 5' CpG islands on the inactive X chromosome. This chromatin structure would then be refractory to the binding of transcriptional activators (such as Sp1 and AP-2) and result in transcriptional silencing of the associated genes. This mechanism is supported by 5-azaC reactivation studies of the human HPRT gene by Sasaki et al. (45). These studies indicate that following hemidemethylation of the HPRT locus on the inactive X chromosome by 5-azaC treatment, a change in chromatin structure of the HPRT gene precedes reactivation and expression of the HPRT gene. This would suggest that DNA methylation may have a role in forming or stabilizing transcriptionally repressed chromatin.

Alternatively, crucial functional sites for DNA methylation may be outside the 5' CpG island and gene and in a region not analyzed by existing studies. However, the ability to reactivate individual X-linked loci by 5-azaC treatment suggests that, although X chromosome inactivation is a chromosomal process, there is likely to be some component of regulation at individual loci.

ACKNOWLEDGMENTS

We thank Kathleen Carlisle for excellent technical assistance.

This work was supported by Public Health Service grant GM44286 from the National Institute of General Medical Sciences to T.P.Y., and a March of Dimes Predoctoral Graduate Research Training Fellowship to I.K.H.

REFERENCES

1. Ben-Hattar, J., P. Beard, and J. Jiricny. 1989. Cytosine methylation in CTF and Sp1 recognition sites of an HSV tk promoter: effects on transcription *in vivo* and on factor binding *in vitro*. *Nucleic Acids Res.* 17:10179-10190.
2. Bird, A. 1992. The essentials of DNA methylation. *Cell* 70:5-8.
3. Bird, A. P. 1986. CpG-rich islands and the function of DNA methylation. *Nature (London)* 321:209-213.
4. Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* 234:47-52.
5. Buschhausen, G., B. Wittig, M. Graessmann, and A. Graessmann. 1987. Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proc. Natl. Acad. Sci. USA* 84:1177-1181.
6. Church, G. M., and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81:1991-1995.
7. Cullen, C. R., P. Hubberman, D. C. Kaslow, and B. R. Migeon. 1986. Comparison of factor IX methylation on human active and inactive X chromosomes: implications for X inactivation and transcription of tissue-specific genes. *EMBO J.* 5:2223-2229.
8. Dracopoli, N. C., W. J. Rettig, A. P. Albino, D. Esposito, N. Archidiacono, M. Rocchi, M. Siniscalco, and L. J. Old. 1985. Genes controlling gp25/30 cell-surface molecules map to chromosomes X and Y and escape X-inactivation. *Am. J. Hum. Genet.* 37:199-207.
9. Ellis, N., E. Keitges, S. M. Gartler, and M. Rocchi. 1987. High-frequency reactivation of X-linked genes in Chinese hamster × human hybrid cells. *Somat. Cell Mol. Genet.* 13:191-204.
10. Fuscoe, J. C., R. G. J. Fenwick, D. H. Ledbetter, and C. T.

- Caskey. 1983. Deletion and amplification of the HGPRT locus in Chinese hamster cells. *Mol. Cell. Biol.* 3:1086-1096.
11. Garrity, P. A., and B. J. Wold. 1992. Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and in vivo footprinting. *Proc. Natl. Acad. Sci. USA* 89:1021-1025.
 12. Gartler, S. M., and B. Burt. 1964. Replication patterns of bovine sex chromosomes in cell culture. *Cytogenetics* 3:135-142.
 13. Gartler, S. M., and A. D. Riggs. 1983. Mammalian X-chromosome inactivation. *Annu. Rev. Genet.* 17:155-190.
 14. Grant, S. G., and V. M. Chapman. 1988. Mechanisms of X-chromosome regulation. *Annu. Rev. Genet.* 22:199-233.
 15. Hansen, R. S., N. A. Ellis, and S. M. Gartler. 1988. Demethylation of specific sites in the 5' region of the inactive X-linked human phosphoglycerate kinase gene correlates with the appearance of nuclease sensitivity and gene expression. *Mol. Cell. Biol.* 8:4692-4699.
 16. Hansen, R. S., and S. M. Gartler. 1990. 5-Azacytidine-induced reactivation of the human X chromosome-linked PGK1 gene is associated with a large region of cytosine demethylation in the 5' CpG island. *Proc. Natl. Acad. Sci. USA* 87:4174-4178.
 17. Harrington, M. A., P. A. Jones, M. Imagawa, and M. Karin. 1988. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc. Natl. Acad. Sci. USA* 85:2066-2070.
 18. Holler, M., G. Westin, J. Jiricny, and W. Schaffner. 1988. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev.* 2:1127-1135.
 19. Hornstra, I. K., and T. P. Yang. 1992. Multiple in vivo footprints are specific to the active allele of the X-linked human hypoxanthine phosphoribosyltransferase gene 5' region: implications for X chromosome inactivation. *Mol. Cell. Biol.* 12:5345-5354.
 20. Hornstra, I. K., and T. P. Yang. 1993. In vivo footprinting and genomic sequencing by ligation-mediated PCR. *Anal. Biochem.* 213:179-193.
 21. Keith, D. H., J. Singer-Sam, and A. D. Riggs. 1986. Active X chromosome DNA is unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine-rich island at the 5' end of the gene for phosphoglycerate kinase. *Mol. Cell. Biol.* 6:4122-4125.
 22. Kerem, B. S., R. Goitein, C. Richler, M. Marcus, and H. Cedar. 1983. In situ nick-translation distinguishes between active and inactive X chromosomes. *Nature (London)* 304:88-90.
 23. Keshet, I., J. Lieman Hurwitz, and H. Cedar. 1986. DNA methylation affects the formation of active chromatin. *Cell* 44:535-543.
 24. Kovesdi, I., R. Reichel, and J. R. Nevins. 1987. Role of an adenovirus E2 promoter binding factor in E1A-mediated coordinate gene control. *Proc. Natl. Acad. Sci. USA* 84:2180-2184.
 25. Levine, A., G. L. Cantoni, and A. Razin. 1992. Methylation in the preinitiation domain suppresses gene transcription by an indirect mechanism. *Proc. Natl. Acad. Sci. USA* 89:10119-10123.
 26. Lewis, J., and A. Bird. 1991. DNA methylation and chromatin structure. *FEBS Lett.* 285:155-159.
 27. Liskay, R. M., and R. J. Evans. 1980. Inactive X chromosome DNA does not function in DNA-mediated cell transformation for the hypoxanthine phosphoribosyltransferase gene. *Proc. Natl. Acad. Sci. USA* 77:4895-4898.
 - 27a. Litt, M. D., I. K. Hornstra, and T. P. Yang. Unpublished results.
 28. Lock, L. F., D. W. Melton, C. T. Caskey, and G. R. Martin. 1986. Methylation of the mouse *Hprt* gene differs on the active and inactive X chromosomes. *Mol. Cell. Biol.* 6:914-924.
 29. Lock, L. F., N. Takagi, and G. R. Martin. 1987. Methylation of the *Hprt* gene on the inactive X occurs after chromosome inactivation. *Cell* 48:39-46.
 30. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
 31. McBurney, M. W. 1988. X chromosome inactivation: a hypothesis. *Bioessays* 9:85-88.
 32. Meehan, R. R., J. D. Lewis, and A. P. Bird. 1992. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* 20:5085-5092.
 33. Meehan, R. R., J. D. Lewis, S. McKay, E. L. Kleiner, and A. P. Bird. 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58:499-507.
 34. Mohandas, T., R. S. Sparkes, and L. J. Shapiro. 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* 211:393-396.
 35. Monaco, A. P., and L. M. Kunkel. 1988. Cloning of the Duchenne/Becker muscular dystrophy locus. *Adv. Hum. Genet.* 17:61-98.
 36. Monk, M. 1990. Changes in DNA methylation during mouse embryonic development in relation to X-chromosome activity and imprinting. *Philos. Trans. R. Soc. Lond. Biol.* 326:299-312.
 37. Mueller, P. R., and B. Wold. 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* 246:780-786.
 38. Ohno, S., W. D. Kaplan, and R. Kinoshita. 1959. Formation of sex chromatin by a single-X chromosome in liver cells of *Rattus norvegicus*. *Exp. Cell Res.* 18:415-418.
 39. Pfeifer, G. P., S. D. Steigerwald, R. S. Hansen, S. M. Gartler, and A. D. Riggs. 1990. Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl. Acad. Sci. USA* 87:8252-8256.
 40. Pfeifer, G. P., S. D. Steigerwald, P. R. Mueller, B. Wold, and A. D. Riggs. 1989. Genomic sequencing and methylation analysis by ligation mediated PCR. *Science* 246:810-813.
 41. Pfeifer, G. P., R. L. Tanguay, S. D. Steigerwald, and A. D. Riggs. 1990. In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. *Genes Dev.* 4:1277-1287.
 42. Razin, A., and H. Cedar. 1991. DNA methylation and gene expression. *Microbiol. Rev.* 55:451-458.
 43. Riordan, J. R., J. M. Rommens, B.-S. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielinski, S. Lok, N. Plavsic, L.-C. Chou, M. L. Drumm, M. C. Iannuzzi, F. S. Collins, and L.-C. Tsui. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245:1066-1080.
 44. Rommens, J. M., M. C. Iannuzzi, B.-S. Kerem, M. L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J. L. Cole, D. Kennedy, N. Hidaka, M. Zsiga, M. Buchwald, J. R. Riordan, L.-C. Tsui, and F. S. Collins. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 245:1059-1065.
 45. Sasaki, T., R. S. Hansen, and S. M. Gartler. 1992. Hemimethylation and hypersensitivity are early events in transcriptional reactivation of human inactive X-linked genes in a hamster $\times$ human somatic cell hybrid. *Mol. Cell. Biol.* 12:3819-3826.
 46. Singer-Sam, J., M. Grant, J. M. LeBon, K. Okuyama, V. Chapman, M. Monk, and A. D. Riggs. 1990. Use of a *HpaII*-polymerase chain reaction assay to study DNA methylation in the *Pgk-1* CpG island of mouse embryos at the time of X-chromosome inactivation. *Mol. Cell. Biol.* 10:4987-4989.
 47. Stout, J. T., and C. T. Caskey. 1985. HPRT: gene structure, expression, and mutation. *Annu. Rev. Genet.* 19:127-148.
 48. Toniolo, D., G. Martini, B. R. Migeon, and R. Dono. 1988. Expression of the G6PD locus on the human X chromosome is associated with demethylation of three CpG islands within 100 kb of DNA. *EMBO J.* 7:401-406.
 49. Venolia, L., and S. M. Gartler. 1983. Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. *Nature (London)* 302:82-83.
 50. Venolia, L., S. M. Gartler, E. R. Wassman, P. Yen, T. Mohandas, and L. J. Shapiro. 1982. Transformation with DNA from 5-azacytidine-reactivated X chromosomes. *Proc. Natl. Acad. Sci. USA* 79:2352-2354.
 51. Verkerk, A. J., M. Pieretti, J. S. Sutcliffe, Y. H. Fu, D. P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. F. Victoria, F. P. Zhang, et al. 1991. Identification of a gene (FMR-1) containing a CGG

- repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**:905-914.
52. **Watt, F., and P. L. Molloy.** 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* **2**:1136-1143.
53. **Wolf, S. F., D. J. Jolly, K. D. Lunnen, T. Friedmann, and B. R. Migeon.** 1984. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proc. Natl. Acad. Sci. USA* **81**:2806-2810.
54. **Yen, P. H., T. Mohandas, and L. J. Shapiro.** 1986. Stability of DNA methylation of the human hypoxanthine phosphoribosyltransferase gene. *Somat. Cell Mol. Genet.* **12**:153-161.
55. **Yen, P. H., P. Patel, A. C. Chinault, T. Mohandas, and L. J. Shapiro.** 1984. Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. *Proc. Natl. Acad. Sci. USA* **81**:1759-1763.