Chromatin Structure and de Novo Methylation of Sperm DNA: Implications for Activation of the Paternal Genome

Mark Groudine and Kathleen F. Conkin

The paternal genome is, according to recent reports, somehow marked for selective expression during early development. For example, the male pronucleus is essential for the formation of the extraembryonic layers during embryogenesis of the mouse (*1*). In addition, the paternal X chromosome is selectively inactivated in the extraembryonic tissues of various species (*2*). Thus, while it is formally possible that paternal DNA is completely reprogrammed on fertilization by components of the egg, there is evidence for the templating of information for differential gene expression by some component of sperm. We have initiated a series of experiments to determine whether, as is the case in other systems (see below), the structure of sperm chromatin or the pattern of methylation of sperm DNA (or both) might provide a basis for the templating of information important in the selective expression of paternal genes early in development.

Changes in chromatin structure are associated with the transcriptional activation of eukaryotic genes. For example, active genes are preferentially sensitive to digestion with deoxyribonuclease I (DNase I) (*3*) and contain sites hypersensitive to digestion with several nucleases including DNase I, nuclease S1 (S1), and restriction endonucleases (*4*). Hypersensitive sites are also present within or near several hormone binding sites and sequences associated with enhancers (*4*). In some cases, these sites appear when the associated gene is transcriptionally activated; in other cases, their appearance precedes the onset of transcription (*4*). In attempting to understand the developmental significance of the appearance of changes in chromatin structure during embryogenesis, we previously reported that sites hypersensitive to DNase I could be propagated continuously through more than 20 cell divisions in the absence of the activity that was initially responsible for generating the altered chromatin structure (*5*). These experiments suggested that at least some hypersensitive sites might serve as structural analogs for the transmission of early developmental signals to progeny cells, perhaps through a mechanism involving protein bound to the DNA sequences comprising the hypersensitive site (*5*, *6*).

Abstract. *The chromatin structure characteristic of constitutively expressed genes, tissue-specific genes, and inactive genes is absent in chicken sperm chromatin. However, point sites of undermethylation in sperm DNA within constitutively expressed genes, but not within globin genes or an inactive gene, correspond to the location of regions of altered chromatin structure (hypersensitive sites) in somatic tissue and spermatogonial cells. A de novo methylation process whereby regions within and flanking these genes become methylated, but which excludes the methylation of sequences within hypersensitive sites, occurs between the spermatogonial stage and the first meiotic prophase. These undermethylated regions may play a role in the activation of the paternal genome during embryogenesis.*

In addition to the structural features of chromatin revealed by nucleases, actively transcribed genes are relatively undermethylated compared to inactive genes (*7*). The idea that DNA methylation might also play a role in the propagation of structural or transcriptional information has been discussed (*7*).

Several groups of investigators have shown that significant changes in structure and expression of chromatin occur during vertebrate spermatogenesis. For example, prior to the condensation of the nucleoprotein complex in mature sperm of various species, protamines replace some classes of histones and transcriptional activity ceases (*8*). In addition, changes in the level of methylation of sperm DNA as compared to somatic DNA have been found for a number of genes in most vertebrates thus far examined (*7*). Thus, these phenomena present an interesting problem. If signals provided by hypersensitive sites or undermethylation of specific regions of DNA are important in providing information for gene expression, how is this information maintained in the hypermethylated, hypercondensed sperm chromatin in a form that might be important in permitting the expression of specific genes on the paternal chromosomes during early development?

As an initial approach to investigate the possibility that information important in determining the developmental activation of specific paternal genes might be encoded in some aspect of sperm chromatin, we analyzed the chromatin structure and methylated state in sperm DNA of constitutively expressed, tissue-specific, and inactive genes of the chicken. We thought that, if chromatin structural elements are involved in the propagation of information from sperm to developing embryo, then housekeeping genes like thymidine kinase (TK), which are transcribed early in spermatogenesis (that is, in spermatogonial cells), would maintain the hypersensitive sites associated with the expression of these genes in the transcriptionally inert mature sperm chromatin, whereas inactive genes such as the avian endogenous virus *ev*-1 and tissue-specific genes like the globin gene, which are not active in spermatogonia, would be propagated in an inactive chromatin structure in sperm nuclei. Our experimental results reveal that the chromatin structure of all genes examined is different in sperm from that in various somatic tissues and spermatogonia. Hence, the maintenance in sperm chromatin of structural features associated with active genes is not involved in templating the paternal genome. However, our results do reveal an unusual difference in the pattern of methylation of

Mark Groudine is an associate member of the Department of Genetics at the Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, and is an associate professor in the Department of Radiation Oncology, University of Washington, School of Medicine, Seattle 98195. Kathleen F. Conklin is a postdoctoral fellow in the Department of Genetics at the Fred Hutchinson Cancer Research Center.

genes in sperm DNA; constitutively expressed genes, but not inactive or tissue-specific genes, contain undermethylated point sites that correspond to the locations of nuclease hypersensitive sites detected in active copies of these genes in spermatogonial and somatic cells. We also find that a de novo methylation process occurs between the spermatogonial and primary spermatocyte stages of spermatogenesis. Since this observed wave of methylation specifically excludes DNA sequences included within nuclease hypersensitive sites in spermatogonial cells, these results are compatible with a model whereby de novo methylation operant during spermatogenesis could play a role in the templating of information in sperm DNA.

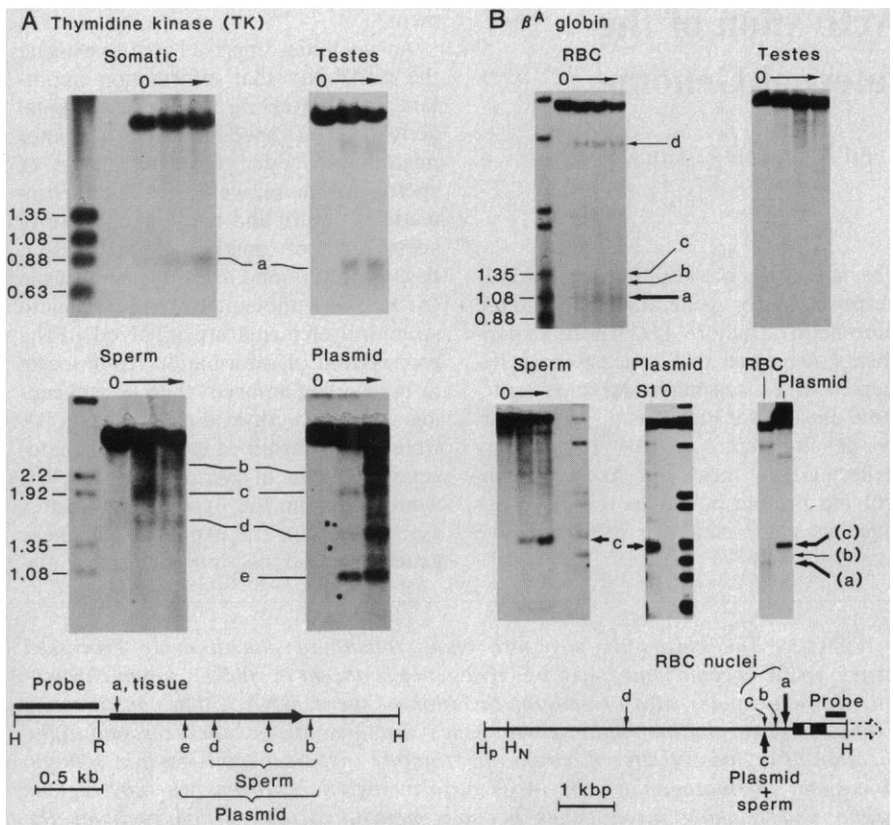


Fig. 1. Hypersensitive sites in sperm chromatin resemble those observed in supercoiled plasmid DNA, rather than those characteristic of active chromatin. Nuclei from designated tissues were digested under previously described conditions (*11, 20*) with increasing concentrations (as designated by the direction of the arrow) of DNase I or S1. DNA was isolated, digested with Hind III (A, B) or Hind III + Eco RI (C), and subjected to blot hybridization (*21*) with ^{32}P-labeled nick-translated probe after electrophoresis in 1 or 1.5 percent agarose gels. Plasmid DNA was digested with S1 (*11*), and processed similarly. Markers consisted of a combination of ^{32}P-labeled Hind III digests of λDNA and Hae III digested ϕX DNA; the weight average molecular weights of the relevant markers are indicated. Sub-bands indicative of hypersensitive sites are indicated by arrows and labeled with letters a through e. The locations of the corresponding hypersensitive sites are shown on the line drawings; H, Hind III; R, Eco RI. (A) Analysis of hypersensitive sites of the thymidine kinase (TK) gene. Results were obtained with a probe spanning the 675-bp region between the H and R sites indicated on the line drawing, where the bold area indicates the coding portion of the TK gene, and the arrow represents the direction of transcription. (B) β^A-globin hypersensitive sites. The end-labeled probe is indicated on the line drawing and consists of a 500-bp Hind III–Bgl II fragment. In the analysis of the plasmid S1 sites, O is the undigested plasmid DNA, S1 is S1-digested plasmid or nuclear DNA. The two 5′ Hind III sites shown on the line drawing, HP and HN, correspond to a sequence polymorphism in the plasmid (HP) compared to genomic/nuclear (HN) DNA. (C) *ev*-1 hypersensitive sites. The end-labeled probe indicated on the line drawing consists of a 1-kbp Hind III–Pvu II fragment. The bold line in the drawing represents proviral sequences, filled-in boxes represent the viral LTR's, the thin line indicates host DNA flanking *ev*-1. The plasmid used for the S1 analysis is defined by the Hind III (H) and Eco RI (R) sites on the line drawing.

Chromatin structure in sperm. In order to determine whether an active chromatin structure in genes destined to be expressed in the early embryo is maintained in mature sperm chromatin, we compared the chromatin structure, as assayed by the pattern of nuclease hypersensitive sites, of the chicken thymidine kinase gene, the adult β^A-globin gene, and an inactive endogenous avian retrovirus, *ev*-1 (*9*), in mature sperm to the structure of these genes in somatic tissues and in the testes of 4-week-old chickens. By histological studies, these early testes contain a minimum of 40 percent replicating spermatogonial cells, but few more mature spermatogenic cells. For all three genes examined (Fig. 1), the pattern of nuclease hypersensitive sites detected in sperm chromatin is distinctly different from those patterns found in chromatin analyzed from cells which express these genes, or from cells in which these genes are inactive. For example, no hypersensitive site is present at the 5′ end of the TK gene in sperm chromatin, whereas in other tissues, including red blood cells (RBC) and the early testes, the 5′ hypersensitive site is the dominant site (Fig. 1A) (*10*). Similarly, our analysis of the location of the hypersensitive sites of the adult β^A-globin gene reveals that in sperm chromatin the dominant site near the 5′ end of β^A (Fig. 1B, site c) is different from the dominant hypersensitive site in the red blood cells in which this gene is expressed (Fig. 1B, site a) (*11*). Our analysis also reveals that the inactive *ev*-1 provirus displays several nuclease hypersensitive sites in sperm chromatin, whereas no such sites are typically evident in any other somatic tissue examined (Fig. 1C) (*12*). It has been shown previously that expression of this inert retrovirus can be induced by treatment of *ev*-1–containing cells with 5-azacytidine; under these conditions, the provirus exhibits hypersensitive sites in positions analogous to those normally found in active proviral DNA's (*12*). Comparison of these azacytidine-induced hypersensitive sites to the sperm-specific hypersensitive sites of *ev*-1 also reveals that the pattern of sperm hypersensitive sites is different from those found in the activated proviral chromatin (*12*).

In attempting to understand the basis of these different hypersensitive sites in sperm chromatin, we considered that

such sites might result from the strain induced by the extreme packaging of DNA during spermatogenesis. It has been shown that supercoiled, but not relaxed, DNA will relieve some of the supercoil-induced strain by forming alternative DNA conformations that are preferentially sensitive to S1 (*11, 13*). Thus, we compared the locations of hypersensitive sites in sperm chromatin to the location of S1 sites in supercoiled plasmid DNA containing genomic inserts of identical sequence. As shown in Fig. 1A, S1 digestion of supercoiled plasmid DNA that contains the chicken TK gene reveals a number of hypersensitive sites, none of which map to the 5′ end of the gene, although several correspond to those sites observed in mature sperm DNA (sites b, c, and d). Similarly, the prominent S1 site in the supercoiled plasmid containing the adult β^A-globin gene is the most 3′ of the cluster of sites at the 5′ end of this gene (Fig. 1B, site c) (*13*), the same site that is dominant in sperm but not RBC chromatin. The difference in location between the major RBC and plasmid or sperm sites is shown in Fig. 1B. In addition, several minor sites are coincident in sperm and plasmid, but are absent in RBC nuclei (Fig. 1B). In the case of *ev*-1, at least three of the nuclease hypersensitive sites in sperm chromatin (Fig. 1C, sites a, b, and c) are coincident with those in plasmid DNA's that contain a portion of this provirus.

Thus, the similar plasmid and sperm S1 sites suggest that the hypersensitive sites that can be assayed in sperm chromatin may be the consequence of the fashion in which DNA is packaged during spermatogenesis. In addition, these results suggest that when sperm nuclei enter the egg, the chromatin structural characteristics of inactive and active genes are not present and therefore cannot provide the proposed structural cues necessary to signal differential gene expression early in embryogenesis.

Methylation of sperm DNA. In order to determine whether genes destined to be active in the early embryo might display a characteristic pattern of methylation of CpG dinucleotides relative to inactive genes or to tissue-specific genes activated later in development, we analyzed the methylated state in sperm DNA of two endogenous avian retroviruses *ev*-1 and *ev*-3 (*9*), the chicken TK gene, and the chicken β-globin gene cluster. The assay for DNA methylation consists of digesting DNA with the restriction enzymes Msp I or Hpa II, both of which cleave at the sequence CCGG. Hpa II will not cut DNA that is methylated at either C (cytosine) residue, whereas Msp I will cleave this sequence regardless of the state of methylation of the C in the CpG (guanine) dinucleotide (*14*). Our former analyses of *ev*-1 and *ev*-3 in somatic tissue revealed that, while the inactive *ev*-1 provirus is highly methylated in all somatic tissues, the active *ev*-3 provirus, which contains more than 20 Hpa II sites detectable by Southern blotting, displays no CpG methylation at any of these Hpa II sites (*12*). As shown in Fig. 2A, Hpa II digestion of DNA obtained from the sperm of an *ev*-1–containing rooster reveals that this provirus is also highly methylated in sperm DNA. However, analyses of proviral DNA in sperm obtained from a rooster containing both *ev*-1 and *ev*-3 reveals a prominent viral related band of 3.7 kilobase pairs (kbp) (Fig. 2, A and B), which is not detected after Hpa II digestion of RBC DNA from the same rooster (Fig. 2B). This 3.7-kbp Hpa II fragment has been observed in the sperm DNA of *ev*-3–containing roosters from several independently derived flocks (*15*). In addition, no such band is observed upon the digestion of either sperm (Fig. 2, A and B) or RBC DNA (Fig. 2B) from any *ev*-1 only roosters, indicating that the 3.7-kbp band is *ev*-3–specific. The 10-kbp fragment (Fig. 2) is also specific to *ev*-3 sperm DNA, but is variably observed, and represents partial cutting at site D shown in the line drawing in Fig. 2. A faint but reproducible 0.25-kbp band is also observed in the *ev*-3 containing DNA from sperm (Fig. 2A). Mapping of these *ev*-3 sperm-specific Hpa II bands by redigestion of the Hpa II samples with either Eco RI, Sst I, or Bam HI (Fig. 2C) reveals that these sites of undermethylation correspond to sites B, C, and D illustrated on line drawing in Fig. 2; the hatched region corresponds to approximately 16 Hpa II sites present in *ev*-3 (*12*). Whereas all of these sites are undermethylated in somatic tissue, only site D is undermethylated in sperm DNA. These results show that the active *ev*-3 provirus, in contrast to the inactive *ev*-1 provirus, displays at least three sites that are undermethylated in sperm.

The possible significance of these undermethylated point sites in *ev*-3 proviral DNA of sperm became apparent when their locations were compared to the location of nuclease hypersensitive sites in these same regions in chromatin of somatic origin. We previously reported that *ev*-3 contained two hypersensitive sites, one in each of the two long terminal repeats (LTR's) of the provirus (*12*); further mapping of the chromatin structure of this active provirus has revealed the presence of three additional hyper-

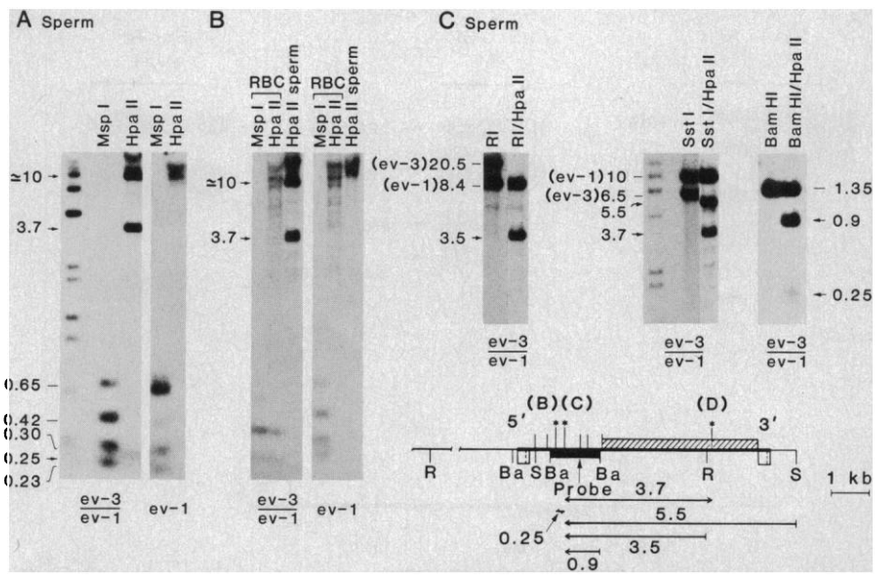


Fig. 2. Undermethylated Hpa II sites in sperm DNA of the active avian endogenous retrovirus *ev*-3. (A and B) DNA from mature sperm and RBC of *ev*-1 or *ev*-1 and *ev*-3 roosters harboring *ev*-3 was isolated, digested with the restriction endonuclease Msp I or Hpa II, and hybridized to the probe shown in D. The arrows indicate Hpa II generated fragments specific to *ev*-3–containing DNA. (C) The same DNA samples described in 2A above were digested with Eco RI (RI), Sst I (S), Bam HI (Ba), Hpa II, or combinations of the above to map the locations of the undermethylated Hpa II sites in sperm DNA. As shown in the drawing, those combinations of digestions reveal sites B, C, and D to be hypomethylated in sperm DNA. The cross-hatched area refers to a region of *ev*-3 containing 16 Hpa II sites. All 16 are undermethylated in somatic tissues (*15*), whereas only site D is undermethylated in sperm DNA. The straight lines on the top of the drawing indicate known Hpa II sites in *ev*-3, all of which are undermethylated in somatic DNA. Derivation of the locations of undermethylated sites B, C, and D in sperm DNA is shown below the line drawing.

sensitive sites. As shown in Fig. 3A, Sst I digestion of DNA from cells that contain *ev*-1 and *ev*-3 reveals bands of 10 and 6.5 kbp corresponding to *ev*-1 and *ev*-3, respectively. Analysis of Sst I digested DNA derived from DNase I or S1 treated nuclei reveals two additional bands of 5.8 and 4.5 kbp. The DNase I or S1 cleavage sites that generate these sub-bands map to the locations indicated on the line drawing at the bottom of the figure, and are derived from *ev*-3 since digestion of nuclei containing only *ev*-1 reveals no such sub-bands (Fig. 1C) (*12*). Additional sub-bands revealed by Eco RI digestion of DNA isolated from DNase I or S1 treated nuclei correspond to hypersensitive sites A, B, and C (Fig. 2, line drawing). Although the data in Fig. 3 are derived from an analysis of RBC nuclei, these hypersensitive sites are detectable in the nuclei of every somatic tissue analyzed thus far, as well as in nuclei from the testes of 4-week-old roosters (*15*). Incubation of intact nuclei with restriction endonucleases has shown that DNA sequences within S1 or DNase I hypersensitive sites are often preferentially accessible to digestion with these nucleases as well (*16*). Using this assay, we have observed that even though all detectable Hpa II sites in DNA purified from *ev*-3–containing cells are undermethylated in RBC, only those sites associated with sequences in the DNase I and S1 hypersensitive regions are accessible to Msp I on digestion of intact nuclei (Fig. 3A).

Comparison of the location of *ev*-3–associated hypersensitive sites in somatic tissue and immature testes and sites of Hpa II undermethylation in sperm DNA reveals that the three undermethylated Hpa II sites in sperm DNA (Fig. 3, sites B, C, and D in the line drawing) are located within regions defined by hypersensitivity to DNase I and S1 and accessibility to Msp I in these other tissues. Because of the lack of CpG's in methylation-sensitive restriction enzyme sites in the LTR's of this endogenous provirus (*17*), we are unable to determine the methylated state of DNA in the regions corresponding to hypersensitive sites A and E located within the LTR's (Fig. 3A). The coincidence of hypersensitive sites B, C, and D (Fig. 3, line drawing) with the undermethylated point sites in sperm DNA is substantiated by the results in Fig. 3B. As shown in Fig. 3B, sperm DNA digested with Hpa II has been redigested with either Eco RI or Bam HI and run on the same gel as DNA isolated from DNase I treated nuclei and digested with the same restriction enzymes. Blotting and hybridization to the viral probe (Figs. 2 and 3A) indicates that the Hpa II–Eco RI generated 3.5-kbp band derived from sperm DNA comigrates with the similarly sized sub-band generated by Eco RI digestion of DNA isolated from DNase I treated RBC nuclei. The larger sub-band detected with Eco RI digestion of the DNase I samples is a hypersensitive site located in the LTR (site A); since no Hpa II site is present in the LTR, no Hpa II fragment corresponding to cleavage at site A is observed. Similarly, Bam HI digestion of the same samples reveals the comigration of the Hpa II generated band from sperm DNA and the smaller of the two DNase I sub-bands which is generated by DNase I cleavage at site C. Both products are seen with DNase I since the probability of DNase I cleavage at two hypersensitive sites on the same molecule is low (*5, 18*). The absence of a Hpa II digestion product corresponding to cleavage at site B is explained by the fact that Hpa II cleaves at sites B and C in the same DNA molecule, resulting in the 0.25-kbp band described in Fig. 3.

Similar analyses were carried out on the chicken TK gene. A summary of our previous analysis of Hpa II site methylation associated with this gene in somatic tissue and sperm DNA (*10*) is shown in Fig. 4A. Of the multiple Hpa II sites associated with the TK gene, only 1 (site h) is methylated in somatic tissue. In contrast, sites f, g, h, and i are methylat-

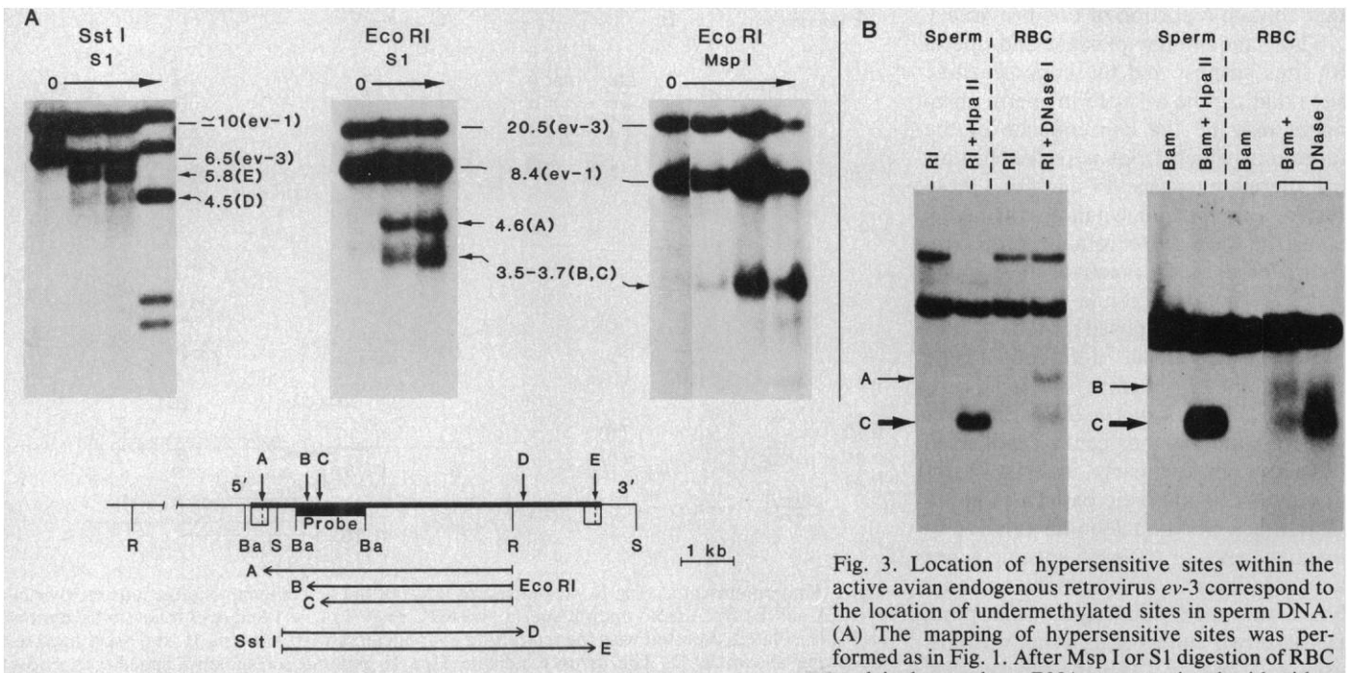
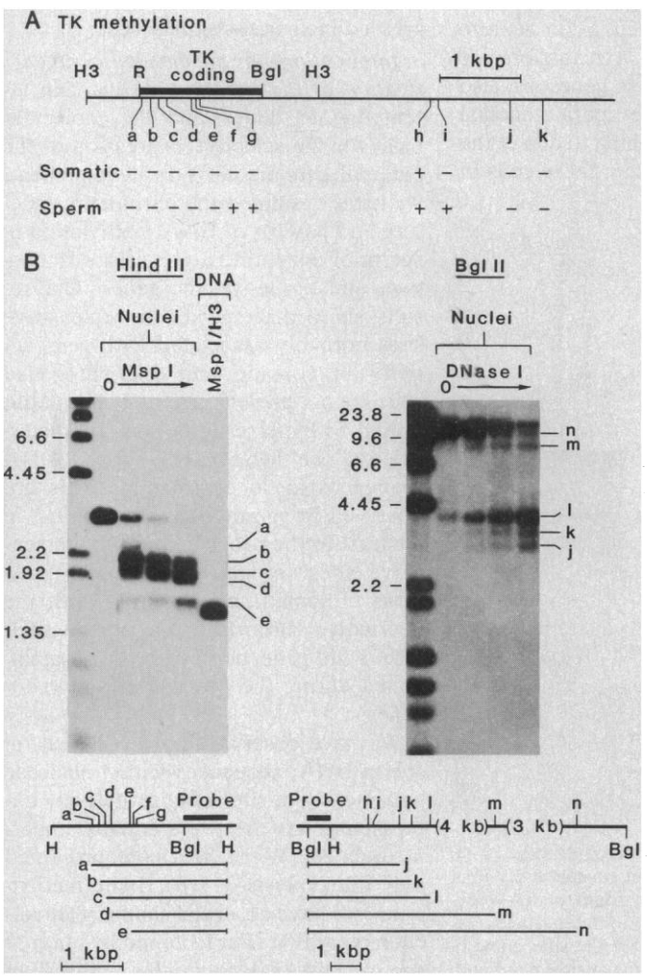


Fig. 3. Location of hypersensitive sites within the active avian endogenous retrovirus *ev*-3 correspond to the location of undermethylated sites in sperm DNA. (A) The mapping of hypersensitive sites was performed as in Fig. 1. After Msp I or S1 digestion of RBC nuclei, the resultant DNA was restricted with either Sst I or Eco RI, and blot hybridized to a 1.35-kb Bam fragment (bold region on the line drawing). The sizes of parent bands are indicated by straight lines, and sub-bands are designated by arrows and corresponding sizes. The locations of hypersensitive sites A through E are indicated on the line drawing, and examples of the mapping of these sites are shown under the line drawing. The open squares indicate the location of the LTR's of *ev*-3; R, Eco RI; S, Sst I; Ba, Bam HI. (B) The identity of the undermethylated point sites in sperm with hypersensitive regions in somatic tissue is confirmed by the comigration of Hpa II generated fragments in sperm DNA with sub-bands generated by DNase I in RBC nuclei. Sperm DNA was digested with either Eco RI or Bam HI and then redigested with Hpa II, revealing the characteristic Hpa II fragments mapped in Fig. 2C. On the same blots, a DNase I digested DNA obtained from DNase I treated RBC nuclei was cleaved with Eco RI or Bam HI, revealing the characteristic sub-bands described in Fig. 3A.

ed in sperm DNA (*10*). If the hypothesis that undermethylated sites in sperm DNA correspond to regions of altered chromatin structure in somatic tissue or spermatogonia, then Hpa II sites a through e and sites j and k should be in regions of altered chromatin structure in these tissues, whereas sites f through i should not. Experimentally, we approached this question by digesting either embryonic RBC, brain, or 4-week chick testes nuclei with Msp I or DNase I and then determining the sites accessible to these endonucleases. As shown for the Msp I treated testicular nuclei (Fig. 4B), Hind III digestion of the purified DNA reveals that sites a through e are accessible to Msp I, whereas sites f and g are inaccessible, as indicated by our failure to observe a fragment corresponding to the limit Msp I–Hind III digestion of naked cellular DNA. By using the same probe and digesting the DNA from the Msp I treated nuclei with Bgl II, we can also conclude that sites h and i are inaccessible to Msp I in these nuclei, whereas sites j and k are accessible to the enzyme (*15*). This conclusion is confirmed by the demonstration that sites j and k correspond to DNase I hypersensitive sites in testicular cells of the 4-week-old chick, whereas sites h and i do not (Fig. 4B). Figure 4B also shows three additional hypersensitive sites (1, m, and n) 3′ to the TK coding region. As determined by Southern blot analysis of partial Hpa II digests of somatic and sperm DNA, there are more than ten Hpa II sites between l and n of which only 1, m, and n are unmethylated in sperm (*15*). Thus, as in the case of *ev-3*, sites of undermethylation in sperm DNA within the TK gene correspond to the location of hypersensitive sites in cells in which this gene is expressed.

In order to ascertain whether the correlation between undermethylated point sites and nuclease hypersensitive sites is restricted to genes marked for expression in the early embryo, we also examined the pattern of methylation in the chicken β-globin gene cluster, which does not become transcriptionally active in erythroid cells until 32 hours of embryonic development (*19, 20*). At that time, the embryonic rho (ρ) and epsilon (ε) β-like globin genes are activated, whereas the β^A- and β^H-globin genes are not transcribed until day 6 to 7 of embryogenesis (*19, 20*). As shown in the line drawing in Fig. 5, Hpa II sites are present within the 5′ and 3′ hypersensitive sites of all four β-like genes (*21*), and are undermethylated in those cells in which the specific gene is transcribed (*22*). No Hpa II cleavage is observed in the 14 kbp of the avian β-globin domain containing sequences 5′ and 3′ to the ρ gene and 5′ to the β^H gene in sperm DNA. This is indicated by the lack of additional bands on Hpa II cleavage of Sst I (S) or Eco RI (R) digested sperm DNA (Fig. 5). Analysis of the β^A and ε gene regions, however, does reveal the generation of an additional band upon Eco RI–Hpa II digestion of sperm DNA (Fig. 5). Extensive mapping of the Hpa II cleavage in this region (*15*) reveals that



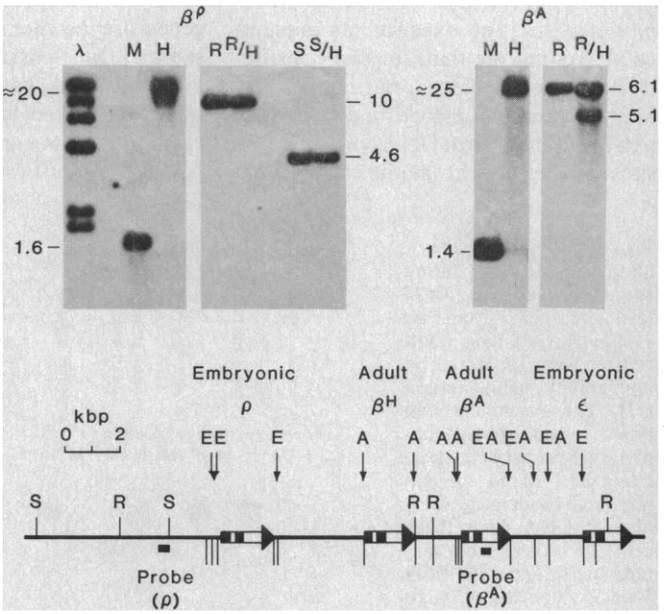


Fig. 4 (left). TK hypersensitive sites in somatic tissue and undermethylated Hpa II sites in sperm DNA are identical in location. (A) Summary of the methylation data in (*10*). The coding portion of TK is defined by the region between the Eco RI (R) and Bgl II (Bgl) sites. Letters a through k on the line drawing denote Hpa II sites. Only site h is methylated in somatic tissue; in contrast, in sperm DNA sites f through i are methylated. (B) Nuclei from 4-week chicken testicular cells were digested with increasing concentrations of Msp I (left) or DNase I (right). The DNA was isolated and restricted with Hind III (left panel) or Bgl II (right panel) and then blot-hybridized to a 700-bp fragment spanning the Bgl II (Bgl) and Hind III (H) sites on the line drawing. Letters a to k correspond to the Hpa II sites depicted in A. The analysis reveals that sites a through e and sites j through n are accessible to Msp I or DNase I in these nuclei, whereas sites f through i are inaccessible. Fig. 5 (right). Methylation in sperm DNA of Hpa II sites within erythroid-specific hypersensitive sites in the chicken β-globin locus. Sperm DNA was digested with Msp I (M), Hpa II (H), Sst I (S), Eco RI (R), or combinations of Hpa II with either Sst I or Eco RI, and then blot hybridized to the β^ρ probe (left panel) or β^A probe (right panel), as indicated on the line drawing below. Locations of relevant restriction sites are S, Sst I; R, Eco RI; H, Hpa II. Hypersensitive sites found in red blood cells are indicated by arrows on the top of the drawing. E, sites found in embryonic RBC; A, sites in adult RBC; EA, sites in both adult and embryonic RBC. Hpa II sites are indicated by unmarked vertical lines.

the undermethylated site is one of three Hpa II sites 5′ to the adult β^A gene, a Hpa II site which is undermethylated in every embryonic and adult tissue examined (*15, 21*). In contrast, all three of these 5′ sites are unmethylated in adult erythroid cells (*22*), and are contained within a red cell–specific hypersensitive site (*16*).

These data reveal that Hpa II sites within hypersensitive sites of the embryonic ρ- and ϵ-globin genes, the adult β^H gene, as well as all but one Hpa II site within the β^A gene are fully methylated in 4-week testicular DNA and sperm DNA. Specifically, Hpa II sites within the 5′ and 3′ hypersensitive sites of the first β-like genes to be activated in the developing embryo are fully methylated in sperm DNA. Thus, in contrast to the two constitutively expressed genes described above, regions of potential regulatory significance within these tissue-specific genes are not marked as different from other regions of the domain in sperm.

Timing of methylation during spermatogenesis. The experiments presented above indicate that, for two constitutively expressed genes, but not for globin, point sites of Hpa II undermethylation in mature sperm DNA correspond to sites of altered chromatin structure in somatic tissues and spermatogonia. These results raise the possibility that these regions of undermethylated DNA may be involved in the templating of information in sperm DNA. If this notion is correct, then the time during development when the methylation pattern of sperm DNA is established could be important in determining which genes will become part of this postulated templating. Thus, we assayed the state of methylation of *ev*-3 and TK in DNA isolated from cells in different stages of spermatogenesis. For this analysis, we compared DNA's from the spermatogonial enriched testes of 4-week-old chickens to DNA isolated from later stages of spermatogenesis including primary spermatocytes in the late pachytene stage of the first meiotic prophase, round spermatids, condensing spermatids, and mature sperm.

In the DNA of early testicular cells, the *ev*-3 and TK genes are as undermethylated as in somatic tissues (Fig. 6), whereas in primary spermatocytes and more mature spermatogenic cells these genes are as methylated as in mature sperm. One possible explanation for these results is that these genes are also hypermethylated in the spermatogonial cells, but that we are failing to detect the contribution of DNA from these cells in our analysis of total testicular DNA. However, since spermatogonia contribute 40 percent of these early testicular cells and, as previously shown, we are able to detect alternative patterns of DNA methylation and chromatin structure contributed by cells making up less than 10 percent of the population under investigation (*20*), we think that this is unlikely. Therefore, we conclude that a de novo methylation of these genes occurs at a developmental time between the transition from spermatogonial cell to primary spermatocyte. In contrast to these constitutive genes, we have observed that the pattern of methylation of the β-globin, ovalbumin, and vitellogenin loci, none of which contain hypersensitive sites in spermatogonial cells (for example, Fig. 1B) is unchanged between spermatogonial cells and mature sperm (*15*). Thus, the observed undermethylation of DNA sequences in the hypersensitive sites of constitutively expressed genes appears to reflect a templating of those genes that contain nuclease hypersensitive sites and are expressed in spermatogonial cells.

Implications for the developmental activation of the paternal genome. In attempting to understand the molecular basis for the selective expression of the paternal genome early in development, we have examined the chromatin structure and pattern of DNA methylation in sperm of constitutively expressed, inactive, and tissue-specific genes. Our results show that specific hypersensitive sites normally associated with gene activity in somatic and spermatogonial cells are not present in sperm chromatin; rather, S1 hypersensitive sites characteristic of the local unwinding of DNA sequences in supercoiled plasmids are observed in sperm chromatin. Thus, in contrast to the propagation of hypersensitive sites through multiple cell divisions in somatic chromatin (*5, 23*), the alternative chromatin structures associated with gene activity are not maintained during the process of spermatogenesis.

We have observed, however, that, in sperm DNA, sequences within nuclease hypersensitive sites of constitutively expressed genes are preferentially undermethylated. While we have presented only the analysis of Hpa II site methylation, we have observed similar phenomenon regarding Hha I site methylation of sperm DNA (*15*). Since the methylation of sequences outside the hypersensitive site regions of these genes occurs between the spermatogonial and primary spermatocyte stages of spermatogenesis,

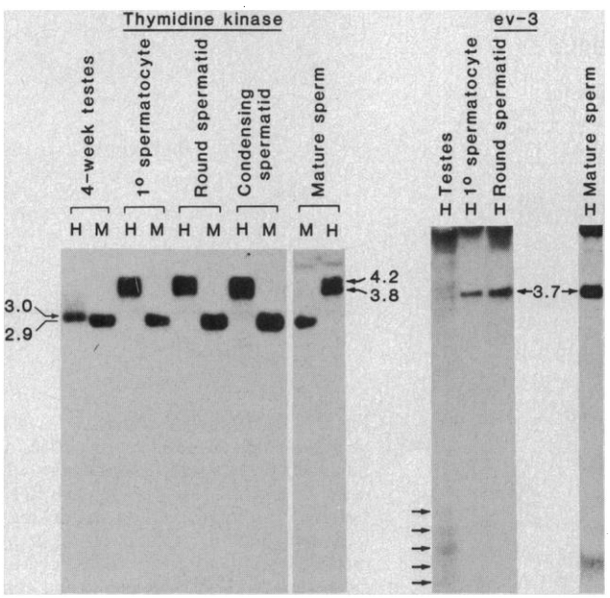


Fig. 6. De novo methylation of TK and *ev*-3 during spermatogenesis. Total testicular cells from an *ev*-1/*ev*-3–containing adult rooster were separated by unit gravity sedimentation (*31*). The prominent cell types isolated by this technique are primary (1°) spermatocytes, round spermatids, and condensing spermatids. DNA from these cells, as well as from mature sperm and cells from 4-week chick testes, were digested with Hpa II (H) or Msp I (M) and blot hybridized to the Hind III–Bgl II TK probe (Figs. 1A and 4B) or the proviral probe (Figs. 2 and 3). For both genes, an increase in methylation is observed between the 4-week testicular DNA and spermatocytes, whereas no further changes in methylation are observed beyond the 1° spermatocyte stage. The arrows indicate the Hpa II generated fragments. For TK, Msp I digestion reveals a 2.9-kbp fragment in DNA from all stages, whereas Hpa II digestion reveals a 3.0-kbp fragment in 4-week testes DNA, and 4.2 and 3.8 kbp fragments in DNA from all cells in later stages of spermatogenesis. The 2.9 kbp Msp I generated fragment corresponds to cutting at sites g and h (Fig. 4A); the 3.0 kbp, to cleavage at g and i; and the 3.8 and 4.2 kbp, to cutting at sites e or f and j and k, respectively. For *ev*-3, in addition to the appearance of the 3.7 kbp *ev*-3 Hpa II fragment from the 1° spermatocyte stage onward, all lower molecular weight fragments except the 0.25 kbp band have disappeared, indicating the methylation of the corresponding sites at this developmental time. The higher molecular weight Hpa II fragments in early testes are derived from the inactive *ev*-1 provirus (Fig. 2, A and B).

our results suggest that the general increase in methylation of sperm DNA compared to somatic DNA is the result of a de novo methylation process operant during spermatogenesis, a process that excludes those DNA sequences within hypersensitive sites in spermatogonial chromatin. Since sequences capable of forming hypersensitive sites in the usually inactive provirus *ev*-1 are methylated in sperm DNA, whereas the same regions in the active *ev*-3 provirus are undermethylated in sperm DNA, the basis for this specificity does not appear to reside simply in DNA sequence per se.

These results raise several questions regarding the mechanism whereby undermethylated point sites are established during spermatogenesis, the generality of the correlation between undermethylated point sites in sperm and regions that include hypersensitive sites in spermatogonial cells, and, finally, the significance of this correlation for regulation of expression of the paternal genome after fertilization. Figure 7 presents a model that combines our results with the idea that the maintenance of hypersensitive sites in an undermethylated state in sperm DNA marks such sites as signals for gene activation early in development.

It is possible that the de novo methylation of DNA during spermatogenesis might play a role in, for example, DNA–protein (*24*) (or protamine) interactions important in condensation of the sperm nucleus. The preferential undermethylation at regions of hypersensitive sites could be based on the presence of a protein at such local alterations in chromatin structure (*5, 6*), thereby inhibiting the procession of the methylase, or based on the lack of specificity of the methylase for regions of altered DNA structure. In the former case, the undermethylated regions would be virtual footprints of such proteins, and if these proteins were lost during the process of spermatogenesis, the undermethylated state of these binding sites might then facilitate either the formation of alternative DNA structures (*25*) or the binding of regulatory proteins to these sites early in development. Such a model would thus predict that the maintenance of these hypersensitive sites in an undermethylated state in sperm DNA would mark such sites as signals for gene activation early in development.

If this model is valid, then the correlation between hypersensitive sites and undermethylated sites in sperm DNA should be found for other constitutively expressed genes. In fact, reports from other investigators concerning the methylated state of constitutively expressed genes suggest that the 5′ regions of such genes are undermethylated in sperm DNA (*26*). While the relation of these reported undermethylated sites in sperm DNA to hypersensitive sites in spermatogonia and somatic tissues has not yet been addressed directly, the location of such undermethylated sites in sperm DNA is coincident with a region often associated with nuclease hypersensitivity. It has also been reported that DNA introduced into mouse embryos prior to their implantation can become highly methylated during embryogenesis (*27*). Thus, it would be interesting to determine whether the relationships among gene expression, chromatin structure, and sites of undermethylation that we have defined during spermatogenesis are evident in this system, or if the proposed templating of the paternal genome is unique.

Since we have shown that in sperm DNA tissue-specific genes do not exhibit undermethylation at regions of potential hypersensitive sites [and similar conclusions can be derived from the work of others (*26, 28*)], our model regarding the templating of developmental cues in sperm DNA cannot account for the activation of tissue-specific genes during embryogenesis. In this model, the activation of tissue-specific genes later in development would be the consequence of other developmental cues, which may be present in the egg or activated through subsequent events.

The wave of de novo methylation occurs during a stage when spermatogonial genes are active in transcription, but prior to the reported transcriptional activation of genes specific to the haploid stages of spermatogenesis such as protamines and a testes-specific α-tubulin (*29*). Our model would therefore predict that, in addition to housekeeping genes involved in DNA replication and other nonspecialized functions, germline-specific genes presumed active in spermatogonial cells would also be templated in sperm DNA. Thus, one possibility would be that the observed marking of sperm DNA could be important not only in the early activation of constitutive genes of the paternal genome but also in the initiation of the germline program, due to the maintenance of signals in the hypersensitive site DNA sequences of spermatogonial (that is, germline) specific genes. In contrast, genes whose products are specific to the later stages of spermatogenesis would not be templated in the fashion described above. In addition, since the

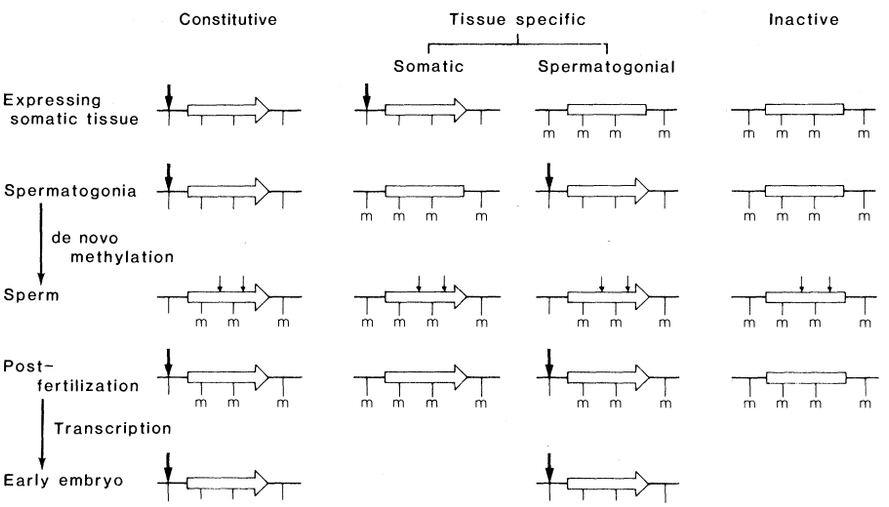


Fig. 7. Model for templating of the paternal genome. In somatic cells, inactive, tissue-specific and constitutive genes display the chromatin structure and pattern of methylation typical of each class; that is, active genes (open boxes with arrowheads) contain hypersensitive sites (thick arrows) and unmethylated CpG dinucleotides (vertical lines), while inactive genes (open boxes without arrowheads) do not contain hypersensitive sites and are methylated (vertical lines marked with an m). These relationships are also evident in the chromatin of spermatogonial cells. During the transition from immature to mature sperm, however, all genes acquire a novel pattern of hypersensitive sites (thin arrows) which more closely resembles that seen in supercoiled plasmid DNA's that contain inserts of these genes. In addition, CpG residues which were included within hypersensitive sites in spermatogonial cells remain specifically undermethylated in mature sperm. These undermethylated point sites may serve to template the paternal genome such that constitutive and spermatogonial-specific genes are activated after fertilization. The methylated sites of constitutive and spermatogonial-specific genes not associated with hypersensitive sites would become progressively unmethylated after continued transcription in the embryo. Activation of tissue-specific genes later in development would require additional events to initiate expression.

de novo methylation process occurs prior to the expression of these postmeiotic genes, increased methylation of these genes might play a role in their transcriptional activation, as has been reported for the mouse transplantation antigen H-2k gene (*30*). These possibilities can be tested by isolating spermatogonial- and postmeiotic-specific genes, and determining the chromatin structure and methylation of these genes during various stages of spermatogenesis, as well as determining the timing of their activation during embryogenesis.

References and Notes

1. M. Surani, S. Barton, M. Norris, *Nature (London)* **308**, 548 (1984); J. McGrath and D. Solter, *Cell* **37**, 179 (1984).
2. J. West, W. Frels, V. Chapman, V. Papaioannou, *Cell* **12**, 873 (1977); N. Takagi, N. Wake, M. Sasaki, *Cytogenet. Cell Genet.* **20**, 240 (1978); M. Harper, M. Foster, M. Monk, *J. Embryol. Exp. Morphol.* **67**, 127 (1982).
3. H. Weintraub and M. Groudine, *Science* **193**, 848 (1976).
4. S. C. Elgin, *Cell* **27**, 413 (1981); K. Conklin and M. Groudine, in *DNA Methylation and Biological Significances*, A. Razin, H. Cedar, A. Riggs, Eds. (Springer-Verlag, New York, 1984), pp. 293–351.
5. M. Groudine and H. Weintraub, *Cell* **30**, 131 (1982).
6. B. Emerson and G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 95 (1984).
7. Reviewed in A. Razin and A. D. Riggs, *Science* **210**, 604 (1980); A. Riggs, in *DNA Methylation and Biological Significances*, A. Razin, H. Cedar, A. Riggs, Eds. (Springer-Verlag, New York, 1984), pp. 269–278.
8. For reviews: A. R. Bellve, in *Oxford Review of Reproductive Biology*, C. Finn, Ed. (Oxford Univ. Press, London, 1979), vol. 1, pp. 159–261; A. R. Bellve and D. A. O'Brian, in *Mechanisms and Control of Animal Fertilization*, J. F. Hartmann, Ed. (Academic Press, New York, 1983), pp. 56–137.
9. S. M. Astrin *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **44**, 1105 (1980); W. S. Hayward, S. B. Baverman, S. M. Astrin, *ibid.*, p. 1111.
10. M. Groudine and C. Casimir, *Nucleic Acids Res.* **12**, 1427 (1984).
11. A. Larsen and H. Weintraub, *Cell* **29**, 609 (1982).
12. M. Groudine, R. Eisenman, H. Weintraub, *Nature (London)* **292**, 311 (1981); K. F. Conklin, J. M. Coffin, H. J. Robinson, M. Groudine, R. Eisenman, *Mol. Cell. Biol.* **2**, 638 (1982).
13. D. Lilley, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6468 (1980); *Nucleic Acids Res.* **9**, 1271 (1981); N. Panayotatos and R. Wells, *Nature (London)* **289**, 466 (1981).
14. C. Waalwijk and R. Flavell, *Nucleic Acids Res.* **5**, 4631 (1978).
15. M. Groudine and K. F. Conklin, unpublished data.
16. J. D. McGhee, W. Wood, M. Dolan, J. D. Engel, G. Felsenfeld, *Cell* **27**, 45 (1981).
17. F. Hishinuma, P. DeBona, S. Astrin, A. Skalka, *ibid.* **23**, 155 (1981).
18. H. Weintraub, A. Larsen, M. Groudine, *ibid.* **24**, 333 (1981).
19. M. Groudine, H. Holtzer, K. Scherrer, A. Therwath, *ibid.* **3**, 243 (1974).
20. M. Groudine and H. Weintraub, *ibid.* **24**, 393 (1981).
21. J. Stalder *et al.*, *ibid.* **19**, 451 (1980).
22. J. McGhee and G. Ginder, *Nature (London)* **280**, 419 (1979); G. Ginder, M. Whitters, K. Kelley, R. Chase, in *Hemoglobin Switching*, G. Stamatoyannopoulos and A. Neinhuis, Eds. (Liss, New York, 1983), pp. 501–510.
23. J. B. E. Burch and H. Weintraub, *Cell* **33**, 65 (1983).
24. E. Fisher and M. Caruthers, *Nucleic Acids Res.* **7**, 401 (1979).
25. M. Behe and G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1619 (1981); M. Behe, S. Zimmerman, G. Felsenfeld, *Nature (London)* **293**, 233 (1981); H. Weintraub, *Cell* **32**, 1191 (1983).
26. R. Stein, N. Sesaky-Gallili, A. Razin, H. Cedar, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2422 (1983); M. Tykocinski and E. Max, *Nucleic Acids Res.* **12**, 4385 (1984).
27. D. Jahner *et al.*, *Nature (London)* **298**, 623 (1982).
28. L. Sanders-Haigh, B. Blanchard-Owens, V. Ingram, in *Hemoglobin Switching*, G. Stamatoyannopoulos and A. Neinhuis, Eds. (Liss, New York, 1983), pp. 39–52.
29. K. Iatrou, A. Spira, G. Dixon, *Dev. Biol.* **64**, 82 (1978); R. J. Distel, K. C. Kleene, N. B. Hecht, *Science* **224**, 68 (1984).
30. K. Tanaka, E. Appella, G. Jay, *Cell* **35**, 457 (1983).
31. A. R. Bellve *et al.*, *J. Cell Biol.* **74**, 68 (1977).

32. We thank Bob Eisenman, Rob Steele, Billy Forrester, Peter Challoner, Mike Goldman, Stan Gartler, and Hal Weintraub for discussion, critical comments; M. Peretz for technical assistance; H. Devitt for typing this manuscript; Hal Weintraub for S1-digested β^{A} plasmid and S1-treated RBC nuclear DNA; H. Robinson for some of the *ev*-3–containing sperm; Ann Skalka for the *ev*-1 plasmid; and Stuart Moss and Tony Bellve for their advice on the separation of chicken testicular cells. Supported by NSF grant PCM 8204696, a scholarship from the Leukemia Society of America (to M.G.), and NIH grant F32CA07476 (to K.F.C.).

7 November 1984; accepted 28 February 1985