Journal of Cellular Biochemistry 44:1–17 (1990)

Structural and Functional Analysis of the Rat Testis-Specific Histone H1t Gene

Sidney R. Grimes*, Steven A. Wolfe, Jeffrey V. Anderson, Gary S. Stein, and Janet L. Stein

Research Service (151), Veterans Administration Medical Center, Shreveport, Louisiana 71101-4295 (S.R.G., S.A.W., J.V.A.); Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, Shreveport, Louisiana 71130-3932 (S.R.G., S.A.W.); Department of Cell Biology, University of Massachusetts Medical Center, Worcester, Massachusetts 01655 (G.S.S., J.L.S.)

A 6.86 kb rat genomic DNA fragment containing the testis-specific histone H1t gene and the histone H4t gene has been sequenced. S1-nuclease protection analyses of total cellular RNA from rat liver and testis showed that histone H1t mRNA was present only in testis. Examination of various highly enriched populations of rat testis cell types revealed that H1t mRNA was found exclusively in a fraction enriched in pachytene spermatocytes. When protein, DNA interactions within the proximal promoter region of the histone H1t gene were examined by electrophoretic mobility shift assays, only minor differences were found in mobility shift patterns of the H1t promoter in assays comparing binding of nuclear proteins from pachytene spermatocytes and early spermatids. However, major differences in binding were observed upon comparing nuclear proteins from rat pachytene spermatocytes to liver. Comparison of binding patterns of rat testis, rat hepatoma H4 cells, HeLa cells, and COS-1 cells also revealed dramatic differences. Transcriptional activity of the histone H1t promoter was examined by measuring H1t promoted chloramphenicol acetyltransferase (CAT) mRNA levels in transient expression assays in transfected rat hepatoma H4 cells, HeLa cells, and COS-1 cells. These assays revealed that the histone H1t promoted CAT gene functioned poorly in HeLa cells and COS-1 cells compared to expression with the parent SV40 promoted vector pSV2CAT. The H1t promoted CAT gene apparently did not work at all in transfected rat hepatoma H4 cells, which is consistent with testis germinal cell specific expression of the histone H1t gene.

Key words: histone genes, gene structure, gene expression, histone mRNA, rat liver, rat testis

Histones are small, highly conserved, basic proteins complexed with DNA in the eukaryotic nucleus [1]. Histone H1, which binds to linker DNA in the region between

*Corresponding author.

Received March 26, 1990; accepted May 8, 1990.

© 1990 Wiley-Liss, Inc.

nucleosomes, is responsible for the higher-order packaging of chromatin and exhibits the most variability in amino acid sequence of any of the histones [2,3]. Typically, there are special sets of histone variants synthesized during different stages of development [4], and there are a number of cases of tissue specific variants [5,6]. The number of histone H1 variants and the relative amount of each variant can differ from tissue to tissue, and for a given tissue can differ from one species to another. Vertebrates possess at least six copies of histone H1 genes [7] in addition to histone H1°, H1t, and H5 genes [3], although estimates range from six to forty copies of each histone gene in mammals.

One of the first germinal histone H1 variants to be identified was the sporocyte specific histone H1 of lilies and tulips [8]. Sperm-specific H1 histones have been reported for sea cucumbers [9], sea urchins [10], and crickets [11]. The group of histones detected in male germinal cells during spermatogenesis in mammals was identified by several laboratories based on unusual electrophoretic mobility in acetic acid-urea gels [12–14]. The testis-specific histone H1t has been found in several species including rat, mouse, hamster, rabbit, bull, monkey, and human [15]. Histone H1t appears to be synthesized only in pachytene spermatocytes and persists through meiosis until all the mammalian histones are replaced by spermatid transition basic proteins [5,16]. Therefore, we have been interested in understanding the mechanisms involved in the tissue-specific expression of the histone H1t gene. Previous reports described the isolation, and subcloning of the rat testis-specific histone H1t gene from a rat genomic library [17–19].

In this paper we report sequencing the entire 6.86 kb EcoRI genomic fragment on which the histone H1t and histone H4t genes reside. We report the mature histone H1t mRNA start and stop sites experimentally determined by S1-nuclease protection analysis, and we demonstrate that steady-state levels of histone H1t mRNA are highest in a cellular fraction highly enriched in pachytene spermatocytes. We also present evidence for the specific binding of testis nuclear proteins to the histone H1t gene promoter by using electrophoretic mobility shift assays. Shift assays show that nuclear proteins from rat hepatoma H4 cells, HeLa cells, and COS-1 cells also bind to the H1t promoter. Transient expression analyses reveal that the histone H1t promoted chloramphenicol acetyltransferase (CAT) gene in the mammalian expression vector pSV2CAT-H1t functions poorly in HeLa cells and COS-1 cells and not at all in H4 hepatoma cells compared to expression with the parent SV40 promoted vector pSV2CAT.

MATERIALS AND METHODS

Reagents and Supplies

The radiolabeled compounds [α^{32}P]-dCTP(3,000 Ci/mmole) and [α^{32}P]-dATP(600 Ci/mmole) were purchased from New England Nuclear. [γ^{32}P]-ATP(4000 Ci/mmole) was ordered from ICN. X-ray film was obtained from Eastman Kodak (XOMAT XAR-5) and DUPONT (CRONEX 7). The pUC18 and pUC19 plasmids and the corresponding replicative form of M13mp18 and M13mp19 bacteriophage DNA were ordered from Bethesda Research Laboratories. The Klenow fragment of DNA polymerase I, T4 DNA ligase, S1-nuclease, and calf alkaline phosphatase were purchased from Boehringer Mannheim Biochemicals. Polynucleotide kinase, T4

DNA polymerase, and Exonuclease III were ordered from New England Biolabs. Restriction endonucleases were obtained from Bethesda Research Laboratories, Boehringer Mannheim Biochemicals, IBI, and New England Biolabs.

Animals and Tissues

Male Sprague-Dawley rats were obtained from Harlan Sprague-Dawley, (Madison, WI) and maintained in a facility fully accredited by AAALAC. All tissues used in these experiments were obtained from animals weighing 250–300 g. Enriched populations of rat testis germinal cell types were prepared by centrifugal elutriation as described [20], with fractions enriched in pachytene spermatocytes and early spermatids being 80% and 70% pure, respectively.

Cell Culture and Transfections

Rat H4 hepatoma cells (H-4-11-E), HeLa cells, and COS-1 cells obtained from the American Type Culture Collection were grown exponentially in minimal essential medium (Eagle) supplemented with 10% fetal bovine serum.

The mammalian expression vector pSV2CAT, obtained from the American Type Culture Collection, and other plasmids were used to transfect cells using calcium phosphate [21]. DNA coprecipitates were left on the cells 16 h before a 3 min glycerol shock. Following the glycerol shock, fresh culture medium was added and the cells were allowed to grow for an additional 48 h. Total cellular RNA was isolated for Northern blots and Hirt extractions were conducted to confirm the presence of the plasmids [22].

Isolation, Subcloning, and Analysis of DNA

Preparation of plasmid DNA and procedures used for electrophoresis of DNA restriction fragments and the Southern blotting technique have been described [18,19]. Probes for these experiments were labeled by oligodeoxynucleotide-primed DNA synthesis using a kit from Pharmacia with [α^{32}P]-dCTP [19]. Hybridization to labeled DNA probes in Southern blot and Northern blot experiments was conducted in 50% formamide solution containing 5 ×SSC (0.75 M NaCl, 0.075 M sodium citrate, pH 7.4), 5 ×Denhardt's (−BSA), 0.1% SDS, and 0.25 mg/ml *Escherichia coli* or salmon sperm DNA at 49°C as described [18]. DNA fragments produced by restriction digestion were electrophoretically separated on low melting agarose gels (FMC BioProducts) and gel slices containing appropriate DNA fragments were melted and ligated to vector DNA using "in-gel" ligations [23]. Small amounts of plasmid DNA were isolated essentially by the protocol of Birnboim and Doly [24]. Large scale plasmid preparations were harvested using alkaline lysis [25] followed by purification on CsCl gradients by centrifugation at 65,000 rpm for 4 h in a Beckman VTi65.2 vertical rotor. The recombinant plasmid pSV2CAT-H1t containing the histone H1t promoter fused to the chloramphenicol acetyltransferase (CAT) gene was constructed as described in the text and in the legend to Figure 6.

Sequence Analysis of DNA

Three DNA fragments from the 6.86 kb EcoRI genomic fragment were subcloned for sequence analysis. The first was a 2.77 kb EcoRI-HindIII fragment, the second was a 1.87 kb HindIII-HindIII fragment, and the third was a 2.23 kb KpnI-

EcoRI fragment. Each fragment was subcloned into the plasmid vector pUC18 or pUC19 for amplification, and the subclones were linearized with an appropriate restriction enzyme and digested with Exonuclease III to produce nested deletions [26]. Shortened DNA fragments were released from the vectors with appropriate restriction enzymes and ligated into bacteriophage M13mp18 or M13mp19. *E. coli* host strain JM101 was transfected by M13 recombinants as described previously [18]. The chain termination method was used to analyze the DNA fragments [18] using cloned T7 DNA Polymerase (Sequenase), available in a kit from the United States Biochemical Corporation. Some DNA fragments were sequenced directly in pUC18 and pUC19 using the Universal and reverse primers from New England Biolabs [27]. Autoradiograms were prepared using Dupont CRONEX 7 film without an intensifying screen and analyzed with the aid of an IBI Gel Reader and the IBI/Pustell DNA Sequence Analysis System on an IBM-PS/2 Model 60 microcomputer. Further analyses were done using the BIONET resource of IntelliGenetics, Inc. (Mountain View, CA).

Isolation and Analysis of RNA

Total cellular RNA samples from rat tissues, from enriched rat testis germinal cell types, and from normal cell culture or transfected cells were isolated and quantitated as described previously [18]. Electrophoretic analyses of total cellular RNA and Northern blot analyses were conducted to determine levels of histone H1t mRNA in various cell types [18] and S1-nuclease protection analyses were performed [18] to find the positions of the 5′ and 3′ ends of the mature mRNAs on the histone H1t gene. In these experiments DNA probes were dephosphorylated and 5′-end-labeled with [γ^{32}P]-ATP using polynucleotide kinase [18]. DNA fragments were 3′ end-labeled by filling recessed ends utilizing T4 DNA polymerase with appropriate labeled and unlabeled deoxynucleotide triphosphates [19].

Nuclear Extracts

Nuclear extracts were prepared from crude rat liver nuclei, from crude rat testis nuclei, and from crude nuclei from enriched rat testis germinal cell types as described [28]. All buffer solutions for extraction contained 0.5 mM phenylmethylsulfonyl fluoride (PMSF), and 5 μg/ml of each of the following four protease inhibitors: pepstatin, leupeptin, chymostatin, and antipain.

Nuclear extracts were also prepared from exponentially growing rat hepatoma H4 cells, HeLa cells, and COS-1 cells as described [29]. All the pelleted nuclear extracts were resuspended in storage buffer, dialyzed against storage buffer for 5 h, and stored at −70°C.

Electrophoretic Mobility Shift Assay

Electrophoretic mobility shift assays (gel retardation assays) were performed essentially as described [30]. Approximately 2 ng of labeled DNA, 1 μg of poly(dI-dC) · poly(dI-dC), and 1–15 μg of nuclear proteins were mixed in binding buffer containing 12 mM Hepes-NaOH, pH 7.9, 12% glycerol, 60 mM KCl, 4 mM Tris-HCl, 0.6 mM EDTA, and 0.6 mM DTT in a final volume of 20 μl. Samples were incubated for 30 min at 4°C and electrophoresed on 4% polyacrylamide gels (80:1 acrylamide: bisacrylamide) using low ionic strength buffer containing 6 mM Tris-HCl, pH 7.9, 3.3 mM sodium acetate, and 1 mM EDTA.

RESULTS

Characterization and Sequencing of the Histone H1t Gene

In this study we have sequenced the 6.86 kb EcoRI rat genomic DNA fragment derived from a 15 kb rat genomic DNA fragment. A partial restriction map of the 6.86 kb EcoRI fragment and the sequencing strategy are shown in Figure 1. The histone H1t and histone H4t genes reside on opposite strands of this genomic fragment separated by 1,300 bp of DNA. The identification of the H1t and H4t genes have been described [18,19]. The protein coding regions are indicated in Figure 1. Unusual elements in the 5′ upstream region of the histone H1t gene, in the region between the histone H1t and histone H4t genes, and in the 3′ region of the 6.86 kb fragment (upstream of the histone H4t gene), respectively, are also indicated in Figure 1. The element upstream from the H1t gene at nucleotide number 602 is $C(A)_4(CA)_{26}$, the element between the H1t and H4t genes at nucleotide 3759 is $(TTTA)_{11}$, and the element upstream from the H4t gene at nucleotide 6695 is $(CA)_9(N)_{17}(T)_{18}GGTTC(T)_7$. These elements are also listed in Table I and shown in Figure 2.

The sequence of the entire 6.86 kb EcoRI genomic fragment containing the histone H1t and histone H4t genes is shown in Figure 2. The entire sequence is also available from GenBank Data Library under Accession Number M28409. The consensus sequence elements near the experimentally determined 5′ mRNA start site (indicated by an arrow at position 2384) are marked. The TATA-box at position 2353, the H1/CCAAT-box at 2330, the GC-box (GGGCGGG) at 2305, and the H1/AC-box (AAACACA) at 2282 have been described [17]. The sequence element CTAGGGAT at position 2320 between the GC-box and the H1/CCAAT-box is an 8 bp element identical to an 8 bp region within a larger promoter element which directs the testis specific expression of the *Drosophila* β2-tubulin gene [31]. In addition, there is a 10 bp oligo(dA) · oligo(dT) tract between the TATA-box and the ATG start codon at position 2411, an element found in the promoter of many eucaryotic genes [32,33]. An inverse complement of the CCAAT-box (ATTGG) is located at position 2422 next to the oligo(dA) · oligo(dT) tract. Furthermore, direct repeats of the H1/AC-box are

TABLE I. Summary of Conserved and Unusual Sequences Within the Promoter of the Testis-Specific Histone H1T Gene*

Proximal promoter region		Downstream sequences		Other genomic sequences	
Position	Sequence element	Position	Sequence element	Position	Sequence element
2282	AAACACA	3109	GGCTCTTTTAAGAGCCACT-TACA	602	$CAAAA(CA)_{26}$
2305	GGGCGGG	3124	mRNA 3′-terminus	3759	$(TTTA)_{11}$
2320	CTAGGGAT			6695	$(CA)_9(N)_{17}(T)_{18}$
2330	ACCAAT				
2342	GCGCC				
2353	TATATAA				
2384	mRNA cap site				
2411	TTTTTTTTTT				
2422	ATTGG				
2453	ATG				

*The numbers indicate the positions of the sequence elements using the numbering system shown in Figure 2.

located at 604 and 5895 and 34 truncated forms of this element (AAACA) or of the inverse complement of this element (TGTTT) are located throughout the 6.86 kb fragment in the noncoding regions upstream and downstream of the H1t and H4t genes. Some of these findings are marked in Figure 2 and summarized in Table I.

Genomic blots of DNA samples isolated from the livers of a male and a female rat indicated that an EcoRI genomic fragment approximately 7 kb in size containing the histone H1t gene was present in female as well as male rats (data not shown).

Testis-Specific Expression of the Histone H1t Gene

The map site of the 3′ terminus of the testis-specific histone H1t mRNA was determined by S1 nuclease protection analysis using a denatured 3′-end labeled probe as described in the legend to Figure 3. Data presented in Figure 3 reveal a protected fragment of 254 nucleotides. The site representing the terminus of this mature somatic histone H1t mRNA is located at the end of the hyphenated dyad symmetry element at nucleotide number 3124 as marked in Figure 2 and as indicated in Table I. Little or no histone H1t mRNA could be detected in rat liver RNA. Since control experiments on the same liver RNA samples reveal the presence of rat histone H4t mRNA [18], the liver RNA appears to be undegraded and we would have detected H1t mRNA if it were present at a significant level.

In an attempt to determine which testis cell types expressed the histone H1t gene, we conducted Northern blot analysis of total cellular RNA from enriched populations of rat testis germinal cell types purified by centrifugal elutriation. The highest steady state level of histone H1t mRNA was found in the cell fraction most enriched in pachytene primary spermatocytes (Fig. 4A, lane 5). Pachytene spermatocytes are premeiotic cells as indicated (Figure 4B). H1t mRNA was not seen in the cell fraction most enriched in postmeiotic early spermatids, although the same quantity of RNA was present in each lane (Fig. 4A, lane 3, and Fig. 4B). Histone H1t mRNA was present in total rat testis RNA samples but at a much lower concentration than in the cell fraction enriched in pachytene spermatocytes (Fig. 4A, lane 6).

Although the histone H1t gene appears to be present in female rats, we have not detected histone H1t mRNA in Northern blots or S1-analyses of total cellular RNA from female rat ovary, liver, or brain (data not shown). In addition, histone H1t protein is not detected in histones derived from these tissues even when assayed in Western blots using specific anti-histone H1t polyclonal antibodies to detect histone H1t (manuscript in preparation).

Nuclear Protein Interactions with the H1t Promoter

Binding of nuclear proteins to DNA within the histone H1t promoter was investigated by electrophoretic mobility shift assay. Although the band designated A is difficult to see, at least seven bands designated A–G are produced when rat testis nuclear proteins are mixed with the histone H1t promoter (Fig. 5A). The 215 bp DNA fragment used in this assay was prepared by filling the Pst I-TthIII 1 fragment as indicated in Figure 2. This fragment includes the consensus sequence elements in the H1t proximal promoter starting at the ATG start codon. In separate experiments (data not shown) there was successful competition with a 100-fold excess of an identical unlabeled DNA fragment. Nuclear proteins from testis cell fractions enriched in both pachytene primary spermatocytes (80%) and early spermatids (70%) produced six of

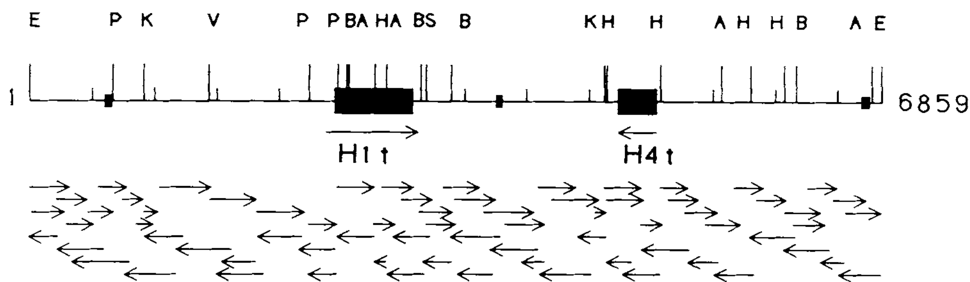

Fig. 1. Restriction map of the rat genomic DNA EcoRI fragment containing the rat testis specific histone H1t gene and the histone H4t gene. The short tic marks represent 500 bp intervals. The restriction enzymes marked on the map include E, EcoRI; P, PstI; K, KpnI; V, PvuII; B, BalI; A, AvaI; H, HindIII; and S, SalI. The coding regions of the H1t gene and the H4t gene are marked as large filled boxes and regions encoding their mature mRNAs are indicated by arrows below the coding regions. The small box near the PstI site upstream from the H1t gene represents a repetitive element $(CA)_{26}$, the small box between the H1t and H4t genes represents a repetitive element $(TTTA)_{11}$, and the small box near the AvaI site upstream from the H4t gene represents a repetitive element $(CA)_9$ and a $(T)_{18}$ element separated by 17 bp. The sequencing strategy for this DNA fragment is indicated by the small arrows at the bottom of the figure.

the seven bands, but scans of autoradiograms revealed that the relative abundance of band C was greater in the cellular fraction enriched in pachytene spermatocytes (lane 3, 8%) than in the fraction enriched in early spermatids (lane 3, 4%), while the abundance of band B was slightly lower in pachytene spermatocytes (lane 3, 5%) when compared to early spermatids (lane 3, 7%) (Fig. 5A). Band D appeared to be absent in pachytene spermatocytes and early spermatids. Further examination revealed that band D, only a minor band in the unfractionated testis cells (lane 3, 4%), was prominant in liver (lane 3, 11%), a tissue type where the H1t gene is not expressed. On the other hand there appeared to be a diminished level of band C in liver (lane 3, 2%) compared to unfractionated testis cells (lane 3, 7%).

Protein-DNA binding was also conducted with nuclear proteins derived from rat hepatoma H4 cells, human HeLa cells, and monkey kidney COS-1 cells. The banding pattern produced by rat hepatoma H4 nuclear proteins contained five of the seven bands (A, C, E–G, Fig. 5B) produced by nuclear proteins from testis (Fig. 5A). In contrast, only two major bands were produced with proteins from HeLa cell and COS-1 cell nuclear extracts. These bands designated C′ and E′ exhibit different mobilities when compared to bands from H4 hepatoma nuclear extracts and testis nuclear extracts. Band E′ has a lower mobility than band E, while band C′ has a greater mobility than band C.

H1t Promoted-CAT Expression

The in vitro transcriptional activity of the histone H1t promoter was examined by measuring the level of chloramphenical acetyltransferase (CAT) mRNA in transient expression assays using the mammalian expression vector pSV2CAT with the H1t promoter fused to the CAT gene as described in the legend to Figure 6 and as shown in Figure 6A. The H1t promoter DNA used in this experiment starts at nucleotide 2452 adjacent to the ATG start codon and extends upstream to nucleotide 2211, as indicated in Figure 2.

A

```
        10         20         30         40         50         60         70
TTCTGAGGTT CCTGCAAGAG AGCCCCAACC TTATACATTC AGTCCTCATC AGTAGTGTTG TTGATATTGA

        80         90        100        110        120        130        140
TGTAAAACAA TAAAGCCTAG ATAAATATTG TAACAAAATG GATAATATTC TTGACCTTCT GATGGTTCTT

       150        160        170        180        190        200        210
GGGACAAAGA ATAATCAGCA CATCTTTTGC ACAAATATTT AGGAACTGTC ACAAACAGCC CTAGTGGGAT

       220        230        240        250        260        270        280
TTCCTTGGTT TCAGATCCTG AGCTCTGGTA ATAAAGGACA AAAGAGGAAA TTCAGGAGAT ATGACTCTGG

       290        300        310        320        330        340        350
AACCTTCACC TCCACCTCAA CTCCAAGATC TAGTTTATGG TTTTCACCAT TTGAAATTCA AGTTTTAATA

       360        370        380        390        400        410        420
TAAACATCTC AAACATAAGT AACAAGACAA CTCACAAACT GCATAAGGAC AAACATGTGA CTTTGAGCAC

       430        440        450        460        470        480        490
AAATGCCAGC CCTCCCCTCT CATCACTGAA CAGTCTCTGA TAAACAACGC TGTGAGTTTA TTCCTAAAAG

       500        510        520        530        540        550        560
GAAATTCTTG AAGAGAAACA AGGAAACAAG TAACATCTCA TAAAATTTCA TTAATTTTGA ATTGCTTTAT

       570        580        590        600        610        620        630
TAAACTTTAA ACATGGTTGA TCAATGTAAG TTCAAAACTC ACAAAACACA CACACACACA CACACACACA

       640        650        660        670        680        690        700
CACACACACA CACACACACA CACACACACC TGCAGTGCTA GAGTTGAATG TAGGCTCTAC CTCATACTAG

       710        720        730        740        750        760        770
ACAGGTGCTT TACTACCAAG ACTCTTCCCC AGTGTATTCT GCATGGTTTT TATTTGGGGA CAGTGTCTTA

       780        790        800        810        820        830        840
TGCAGTTTAT TTACTCTATA GCTTAGGTTT GCTTTGAACT TGGATCATCT CTCAGGTAGC TAGGGTTACA

       850        860        870        880        890        900        910
GTTCGGCACA CCGGGCCTAG TCAATTAACT GTTAAGACAA AAGTACAAAT GTTTTCCTGA GACATAGGTA

       920        930        940        950        960        970        980
CCTACCCACC CTCATGTTTT TCTTTTTCTT TTTAAAAATT TTTATTAGAT ATATTTCTTT ACTTACATAT

       990       1000       1010       1020       1030       1040       1050
CAAATGTTAT TCCCTTTCCC GGTTTCCTGT CCATAAGCCC TCATCCTCTA CCCTCCCCCT CCCCTATATG

      1060       1070       1080       1090       1100       1110       1120
GGTATTCCCC CCATACTTCC CCCTTACCAC CACCCCCCCA GTTTTTCTTA ATAAACAACT TTTCCTCATA

      1130       1140       1150       1160       1170       1180       1190
TTATCTTGAA CTTTTCTGTA ACATAGAGAT CTCTGCCTCA CAAGGGTAGT TATCTCCCAA AAGAACTCGG

      1200       1210       1220       1230       1240       1250       1260
ACTGTATGAC TGCTGCATTG GACTTTATGC TCTACTTCAG TTGATATAAT CACACTACCC CATCTTAGCA
```

B

```
      1270       1280       1290       1300       1310       1320       1330
TTCTCCATAA TGAACTAGCT TAATTAGGCC TCCCTTCTAC TTGAGTATCA TGTTCCCTCC TTTTTTTCTT

      1340       1350       1360       1370       1380       1390       1400
CTTCTTTTGA CATGTTTTCA TTTTCTAGCT AAAGGTGACC TCAAAACTCA CAGCAATCCT CCCTTCTTCA

      1410       1420       1430       1440       1450       1460       1470
TGCTGGGATT GCCAGCAAGT TGCCACCATG CCCAGCTGTA AGCACTCTTA TTTCTAGACA TGTACCATGG

      1480       1490       1500       1510       1520       1530       1540
ATAAAGCCCT GGAAAGAACA TTTCTCTAAC ACATATGTAG ACATGTTTAC CAGCCCAGCC ATATTAAGAG

      1550       1560       1570       1580       1590       1600       1610
ATTGTGAAAT CTGATTCCAG CTTTATAATT TTTAAATTAA GGGTAGGCAC TGCATCTGTT CATCTAGAGG

      1620       1630       1640       1650       1660       1670       1680
TAGATACGAC ATTGGAAGCT GTTGTAATGT GGGATGATGT CACAGGAAGT GGACTATTGT GGTGTTCAGT

      1690       1700       1710       1720       1730       1740       1750
TGGAAAGGAG TACCAAACTC CTTTGATAAC AGCCTTGAGG CTGGATGATT TGACTATATA GATGTTATTG

      1760       1770       1780       1790       1800       1810       1820
GTAAGATGCT TATAACCGCA TCCAAAGCCC TTTTCTTAAG ATTCAGGGGA AACAGCGTGC CTCTCACACG

      1830       1840       1850       1860       1870       1880       1890
ATCTGTGGAG TGTTTATTAG GCCATTCTCC CTTCAAGAAG ACCACACCTT GGCGTTTTGG GATAATTCTA

      1900       1910       1920       1930       1940       1950       1960
GGAACACTGT CTAATTCCAC AGTGAGATAG TCAAGACATC TGTGTTACTT CTGTCAGGTA TATATATAAT

      1970       1980       1990       2000       2010       2020       2030
TTCAGTGAGA ATTAGGGCCT GCATAATTGA TTTTCTTCAC TAGTTAGAAG CCACATATGT AGAAGACCCC

      2040       2050       2060       2070       2080       2090       2100
TCCCAAGATG TTGGATACAT TTTCTACTAA AAGTTACTTT TGAATTAGCA AGAAAAAGAG AAAGTCTTCT

      2110       2120       2130       2140       2150       2160       2170
TCTTCTTTTT GAATTTGATT GAAAATCAAG GGCGTTTCCC ACTAGTTCTG ACATGCAGGA ACATTTGTCA

      2180       2190       2200       2210       2220       2230       2240
ATAAAGGGTA AAAAGTCAGG TGGAATCCAA GTTCCCGTGG GTCACCTGTG TCATAACCTG AGCGATTCTG

  |    2250       2260       2270       2280       2290       2300       2310
CAGAGAGCAC ACATGCAAGT TTCTACTTTT TGGGGGAAAA GAAACACAAA TGCCCCTTCC CCAGGGGCGG

      2320       2330       2340       2350       2360       2370       2380
GGGAGGCGCC TAGGGATGCA CCAATCACAG CGCGCCCTGC TCTATATAAG GCCCCCCCCC GGACCCCCTC

  |>
  |    2390       2400       2410       2420       2430       2440       2450
TCTACTCCAG CGCTGTTCGC TCTCCTGTAC TTTTTTTTTT AATTGGGTGT AGTTCTTGAC CAGCTCTTGA

  |-- 2460       2470     |  2480       2490       2500       2510       2520
CTATGTCGGA AACGGCTCCT GCAGCCTCAA GTACTCTTGT TCCAGCTCCT GTAGAGAAAC CTGCAACTAA
  MET ->                    HIT CODING REGION
```

C

```
      2530       2540       2550       2560       2570       2580       2590
AAGGCGAGGG AAGAAGCCTG GCATGGCCAC TGCTCGCAAA CCTCGGGGTT TCTCGGTTTC CAAGTTGATT

      2600       2610       2620       2630       2640       2650       2660
CCTGAGGCCC TTTCCATGTC TCAGGAACGG GCAGGAATGT CCCTTGCTGC CCTGAAGAAA GCCCTGGCTG

      2670       2680       2690       2700       2710       2720       2730
CGGCTGGCTA TGACGTGGAG AAGAACAACA GTCGTATCAA GCTGGCCCTC AAGAGACTTG TGAATAAGGG

      2740       2750       2760       2770    |  2780       2790       2800
AGTCCTGGTG CAGACCAAGG GCACCGGAGC CTCAGGCTCC TTCAAGCTTA GCAAGAAGGC AGCTTCAGGC

      2810       2820       2830       2840       2850       2860       2870
AACGACAAGG GCAAGGGCAA GAAATCTGCT TCTGCCAAGG CTAAGAAACT GGGCTTGTCC AGGGCCTCGA

|     2880       2890       2900       2910       2920       2930       2940
GATCCCCCAA GAGTAGTAAG ACCAAGGTTG TCAAGAAGCC AAAAGCTACG CCCACAAAGG GTTCTGGGAG

      2950       2960       2970       2980       2990       3000       3010
CAGAAGGAAG ACCAAAGGGG CCAAGGGCTT GCAACAGCGC AAAAGCCCCG CCAAAGCGAG GGCAACCAAC

      3020       3030       3040       3050       3060       3070     ---3080
TCCAACTCTG GGAAGTCAAA GATGGTCATG CAGAAGACCG ACCTAAGGAA GGCAGCAGGA AGAAAGTGAG
                                                                    LYS***

      3090       3100       3110       3120   -->  3130       3140       3150
TTTCAAAGCC AGTTTTCAAA AACCCAAAGG CTCTTTTAAG AGCCACTTAC ATACTTCTTA AAATGGCCAA

      3160       3170       |          3190       3200       3210       3220
ACACTGAGCA AAAGTTAGAG GTGGGCAGTC ACTTAGGTCG ACCTGCTGTT TAAACCCTAG AGTGGAAACC

      3230       3240       3250       3260       3270       3280       3290
TAGGAGGTCC CTATGTGTAG TTTATTTGGT TGTGGCTTTA CATTAGAGTA AAGGTCACCA TTGTAACATG

      3300       3310       3320       3330       3340       3350       3360
TCTTTCATTC TAGAATAAAG ATCACCATTA AAATGTAAAA TGGATGCATT TCATCCTAGA AATGCCACCG

      3370       3380       3390       3400       3410       3420       3430
TATCAAAAAG AGATGACCTC AGTAGTAGCT GGCCACTCTG ATCAAATACT GAACTTAAAT TTACTCATTA

      3440       3450       3460       3470       3480       3490       3500
GAACACCCGT GACTGGCGTC CAGGCTTTCA TAGCACCACT GTGAACCCAG ATATTCAGTG ACTAAGCTAG

      3510       3520       3530       3540       3550       3560       3570
AGTTTATACA TTTTTCTTGA GTTGGATTTT TATTTCCTGA GTGATGGATT AAAAAAATAC AACAGTTCCT

      3580       3590       3600       3610       3620       3630       3640
AAGAAACTAA AATGTCTTTG AAGTTATTGC AATTGTGCGA TGTGTCAACA AAGATTCATT TTTTGTTCCA

      3650       3660       3670       3680       3690       3700       3710
CTTAGAAAGT AGTTTAATAC TATTGATCTA ACAGTTATTG TTGAAATGTA TTGTGTATTT TATGAAAATA
```

D

```
      3720       3730       3740       3750       3760       3770       3780
GCATGTAGGC TTCATGAAAC TGACAAGAGC CATATTAAAA ATTAAGGATT TATTTATTTA TTTATTTATT

      3790       3800       3810       3820       3830       3840       3850
TATTTATTTA TTTATTTATT TATTATGAAA ATGTTAGGGT TTGAGTTCAA CGCTCCAACA TAACAAAAAC

      3860       3870       3880       3890       3900       3910       3920
AGTACCGGGA AAGTCTGGGA TGAAGAATGT AAACAACTCA GTTTTTAAAG TCATTTAACA GTTTTATGGA

      3930       3940       3950       3960       3970       3980       3990
TTAAAGTGGG CCACAAATAT TAAATTTACA TTTATTTTAG GCCTTTGTAT GAGGAGTGAC AAAAGATCAT

      4000       4010       4020       4030       4040       4050       4060
TTAGCAGGGT GGCTTGAGAA ATTGTTCTGT CAAGAATACT TGATGCCTTT AAAGAAGTTT GTATATCAGA

      4070       4080       4090       4100       4110       4120       4130
ACACACAAGT GGCTCACAAC TGCTAACTCC AGCTTCCTGA AAATATGATT CTCGCTTCAG GCAGAGGGCA

      4140       4150       4160       4170       4180       4190       4200
AACACGGCAC ACCCGCACGA ACTTAAAATA TCAAGCCATT GGACAACCAT GCAGTTTGTG CATAAACACT

      4210       4220       4230       4240       4250       4260       4270
ATCAACAGGC GGTCCCTAAC ACTTTTGTGT ATTTAGGCCT TATTAGTTGT ATACAATGAG TAACAGTTAA

      4280       4290       4300       4310       4320       4330       4340
ACATGTCTTC GGTTTTTGTA GTTAGCCCTT GCCTTAGACA AAATGGCTCC TGCTCTTTTG AGCTGGCAAT

      4350       4360       4370       4380       4390       4400       4410
TTGGATGGCG CAAGTTTAGA GACAGGTACG TGTAAGTAAA TTGACTTAAG GTTATTTACA AGGAAACACT

      4420       4430       4440       4450       4460       4470       4480
GAGAAACTGG GCTCACTTAT TAGAAGGGAG CCTTTCAAAG GTAGTCACTT AGCATTGTTA ATACTTCAGT

      4490       4500       4510       4520       4530       4540       4550
TCAGGACTAT CTACTCAAAA CCTTTAGTTA ATGATACAGG GTATATAAGA AAGCTGCATG GTTATGGATT

      4560       4570       4580       4590       4600       4610       4620
AGAATTACAG TCTTACTTTC TAGCTGTGAA CATCCAACCA GAGTATAGCA TAGGACACCG CTTATATGGT
                                                                    <-
      4630       4640       4650       4660       4670       4680       4690
ACCTCGCATG CTCCAGAGCA AAGAAGCTTC GAAAAGTGAA CAACCCTTTG AATGAGTTCG TGGGTGGCCC

      4700       4710       4720       4730       4740       4750       4760
TGAAAAGGGC CTTTTGCAAC AGCTAAGGAA AGCGCATTAG CAGTTTAACC ACCGAAGCCG TAGAGTGTGC
                                                  ***
      4770       4780       4790       4800       4810       4820       4830
GGCCCTGGCG TTTAAGCGCG TAGACCACGT CCATGGCAGT GACCGTCTTG CGCTTGGCGT GCTCCGTGTA

      4840       4850       4860       4870       4880       4890       4900
GGTGACGGCG TCGCGGATCA CGTTCTCCAG GAACACCTTC AGCACACCGC GGGTCTCCTC GTAGATGAGG
```

Figure 2A–D

E

```
      4910       4920       4930       4940       4950       4960       4970
CCGGAGATGC GCTTCACTCC TCCGCGCCGG GCCAGGCGGC GGATGGCGGG CTTTGTGATG CCCTGGATGT
                           <- H4t CODING REGION
      4980       4990       5000       5010       5020       5030       5040
TGTCGCGCAG GACTTTGCGA TGACGCTTAG CGCCACCTTT CCCAAGCCCT TTACCACCTT TACCACGTCC

      5050       5060       5070       5080       5090       5100       5110
AGACATGGTT GTGAGTAAAA GTCAAGGGAA GAAAGCAAGC TTCCCTTTAA ATAGACCTTC TGCGGACCTG

      5120       5130       5140       5150       5160       5170       5180
ATTGAAGACC ACAGGCGTTA GGAGGGGTTT TCGAATCAGG TCCTTGAAAG GGGGTGGGGC AACAAGGATG

      5190       5200       5210       5220       5230       5240       5250
AGTCATTGCA ATCTGAATGT TTTCATTGGC TAGCAAACCT CAGCTATTCT TTGCGCGCTT TGTCTCCTAA

      5260       5270       5280       5290       5300       5310       5320
AGACAGACCA GGTTTAAGTC CTGCCAAAAC CTCATTGTCA TCCCCAACCC CTTCTTTCCC GCGAAAGGCT

      5330       5340       5350       5360       5370       5380       5390
TATTGGCAGA AAAGCAGATA TCCCCATTTC TCGTTCGCTT TTACTTCTAT CGTGTTATTA AATTAATTGA

      5400       5410       5420       5430       5440       5450       5460
TGGTTTCTTT GTGAGGACTC AGGTAAGATG TCCAAAACAG AGTGAATTAA ATAAACCAAG GATGTCCCTT

      5470       5480       5490       5500       5510       5520       5530
GGCAGTAATG TTTAAGCCCG TCCCTGGCAG TTTTGTTTGG GAAGGCAGAA ACATTGAAAA GGCATCTCCA

      5540       5550       5560       5570       5580       5590       5600
ATTGAGCACC CTAACTGCAA ACTTCATATT CCCTATCTCG AGTCAGGGAC CTAATGACTC TATCTTCTTT

      5610       5620       5630       5640       5650       5660       5670
TTTATAAAAA CTGAGACAGT CAAATAAAGT TAGGCTACCG GTTATCGCAA CCTGCTGTGG TTCCAAGTTT

      5680       5690       5700       5710       5720       5730       5740
GTTTCCCTTC CAGCCCTTAT TAAGCACAGC ATGGCTTGCT CAACGTTAAA ACGTGTTTTT TCGAGCTGGC

      5750       5760       5770       5780       5790       5800       5810
AATTACAGTA CAGGGCTAAT ACTTGATTAA GGTCATTCCA AAGCTGACAA AGTTCTTGTG AAGCTTGTGA

      5820       5830       5840       5850       5860       5870       5880
TAAGATTAGT GATGTCAAAA CAAGGCTAAT CCTTTTTTAA AAAGATTTAT CTATCAAAAA GATAAGTACA

      5890       5900       5910       5920       5930       5940       5950
CCATAGCTGT CTTCAAACAC ACCAGAAGTG GGCATCAGAC CTCATTACAA ATGGTTGTGA GTCATCATGT

      5960       5970       5980       5990       6000       6010       6020
GGTTGCTGGG ATTTGAACTC AGGACCTGTA GAAGAACAGT TAGCACTCTT AACCACTGAG CCATCTCTCC

      6030       6040       6050       6060       6070       6080       6090
AGCTCCAGGG TTACAATTCT TATTCAGGCA AATAGGGAGG GAAGTAGAGA AGCTTTTCCT CAAAACACTA
```

F

```
      6100       6110       6120       6130       6140       6150       6160
TTTTTCTTTT TCTCCACTTT GGACAGTTAA AAACCAGCAC CCCTGAACCA TCAGAATGTA TGCTGGGCAG

      6170       6180       6190       6200       6210       6220       6230
GGCATGGCCA AGTGCTGTTT AGTAATATGT AAACAAATGC ATTGGTGTAC TGTATGTGAA AGAGTAGGGA

      6240       6250       6260       6270       6280       6290       6300
TTAGTCACTA AAGAACAAAC AAGTTATGCT CCAAAATGTT GAGGAAGATT TGAAATTTCA GGTGTCAGTC

      6310       6320       6330       6340       6350       6360       6370
AATACACACA CCTTTCATGC TTAAAGCTCT TCAGAATGGC TTGATAGCCA TTATTAAATC AGAGTATTTT

      6380       6390       6400       6410       6420       6430       6440
AAAAAAGATA TTGGTTTAAA AATAACATGA GAAAATTTGG CACCTTACGT GCAACATTTT TTTCTTTGAA

      6450       6460       6470       6480       6490       6500       6510
ATGAGTCTAG GGGTTGCTGC ATTGGAGTTG TCTCTCTTTT TTCCTTGCAC CACTTAAAAG GAAATGGTTC

      6520       6530       6540       6550       6560       6570       6580
CTGTGAGGAC CTCCCAGCAG GCATGAGTCA CTGTTGAAGG AGCCTCAAAG ATCATTGTCT TCTCTACCTT

      6590       6600       6610       6620       6630       6640       6650
CTGTTCTTTT GGCAGCTTGG GAACTCAGCC TTGGCTTAGC ATACCAGCCT TTCCAGTAGC TAGTTCGTAA

      6660       6670       6680       6690       6700       6710       6720
TTCTCTTTTA CTTATGCACT TTCTCTTTGA AAGAAATAGA AATACACACA CACACACACA CAAACTTTGT

      6730       6740       6750       6760       6770       6780       6790
TATTTGTTCT TTTTTTTTTT TTTTTTTGGT TCTTTTTTTC GGAGCTGGGG ACCGAACCCG GGGCCTCGTC

      6800       6810       6820       6830       6840       6850      6859
GTTCCTAGGT AAGCGCTCTA CCACTGAGCT AAATCCCCAG CCCCGTTATT TGTTCTTTTA AGAGAATTC
```

Fig. 2. Nucleic acid sequence of the 6859 bp EcoRI genomic fragment containing the histone H1t and histone H4t genes. The consensus elements underlined in the proximal promoter region beginning at the ATG start codon and moving in an upstream direction include an inverted CCAAT-box (ATTGG), a putative T-box located in an unusual position between the mRNA cap site and the ATG start codon, the TATA-box at 2353, the H1/CCAAT-box at 2330, an 8 bp element at 2320 identical to a portion of an element necessary for testis specific expression within the *Drosophila* β2-tubulin gene, the GC-box (GGGCGG) at 2305, and the histone H1-specific element H1/AC-box (AAACACA) at 2282. The experimentally determined start site and terminus of the mature histone H1t mRNA are indicated by arrows at 2384 and 3124. Note the large number of repeats of the shortened form of the AC-box (AAACA) and of the inverse of this element (TGTTT) scattered upstream of each of the two genes and between the two genes.

Relative in vitro transcriptional efficiencies of the CAT gene in transfected cells were estimated by transient expression assays using Northern blot analysis to estimate CAT mRNA levels. Hirt extracts [22] revealed successful transfection of rat hepatoma H4 cells, HeLa cell, and COS-1 cell (data not shown). Control experiments were conducted with pSV2CAT-H1tR, the same expression vector shown in Figure 6A but with the H1t promoter in the reverse orientation, and with pSV2CAT-Null, identical to pSV2CAT but with the SV40 promoter deleted. When transfection was conducted with pSV2CAT-H1tR, no CAT mRNA could be detected in Northern blots (Fig. 6B, lane 2), while a low level of CAT mRNA was detected with pSV2CAT-H1t (Fig. 6B, lane 1) compared to cells transfected with pSV2CAT (Fig. 6B, lane 6). These experiments revealed successful transfection of COS-1 cells with the expression vectors and that expression of CAT mRNA was high with the parent expression vector pSV2CAT. No measurable CAT mRNA was produced with negative control vectors pSV2CAT-H1tR or pSV2CAT-Null. Low levels of CAT mRNA produced in COS-1 cells transfected with pSV2CAT-H1t were not unexpected given the high degree of sequence homology of the histone H1t promoter to other histone H1 promoters. CAT

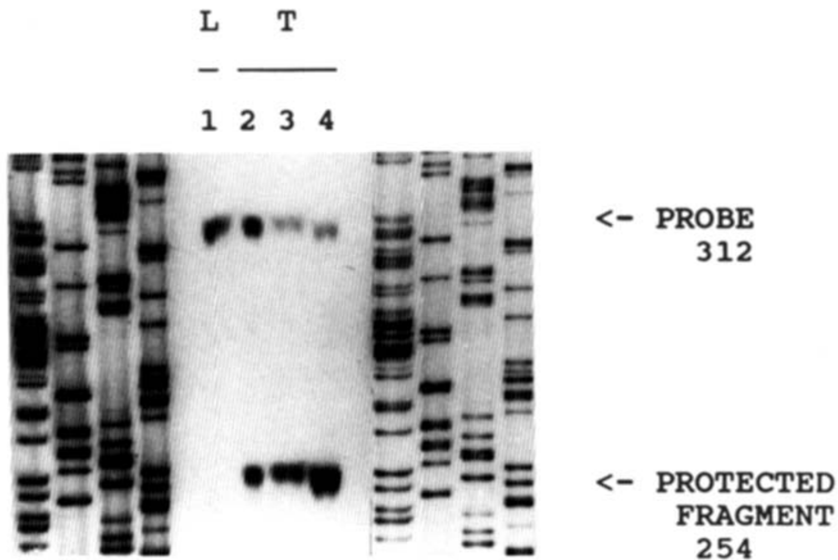

Fig. 3. Determination of the mature histone H1t mRNA terminus by S1-nuclease protection analysis. An MboI-MaeIII DNA fragment spanning the 3′-half of the coding region and past the dyad symmetry element (nucleotides 2871–3182 in Fig. 2) was filled and 3′ end-labeled, denatured, and the 312 base probe was used in S1-nuclease protection analysis of total cellular RNA from rat testis and rat liver. No protected fragment was seen with 12 μg of rat liver RNA (**lane 1**), but a 254 nucleotide protected fragment was seen with three different levels of rat testis RNA (8, 18, and 22 μg in **lanes 2, 3,** and **4,** respectively). The size of the protected fragment places the mature H1t mRNA terminus at the end of the dyad symmetry element at nucleotide 3124 as shown in Table I.

mRNA could not be detected at all in transfected rat hepatoma H4 cells as seen in Figure 6C, lanes 7 and 8. Only low levels of CAT mRNA were measured in COS-1 cells (Fig. 6C, lanes 1 and 2) and HeLa cells transfected with pSV2CAT-H1t (Fig. 6C, lanes 3 and 4). For example, CAT mRNA was detectable in HeLa cells only when 50 μg of total cellular RNA was used in the Northerns (Fig. 6C, lane 4). Multiple bands frequently seen in higher resolution gels in these assays using double stranded DNA probes cut from the CAT coding region indicate that all bands represent mRNA from the CAT coding region.

DISCUSSION

Studies on mechanisms responsible for tissue-specific expression of the rat testis histone H1t gene were initiated by isolating, cloning, and sequencing a 6.86 kb genomic fragment containing the gene. The nucleotide sequence reveals that the fragment contains both the rat histone H1t and H4t genes. The entire sequence is also available from GenBank as described in the Results. Previous examination of the H1t transcriptional start site by S1-nuclease protection studies [18] revealed that the start site was located at nucleotide 2453, which is 69 bp upstream from the ATG initiation codon (Fig. 2). The 3′ mRNA terminus, determined by S1-nuclease protection analysis in this study, is located 45 bp downstream from the termination codon just past the dyad symmetry element at nucleotide 3124, as indicated in Table I.

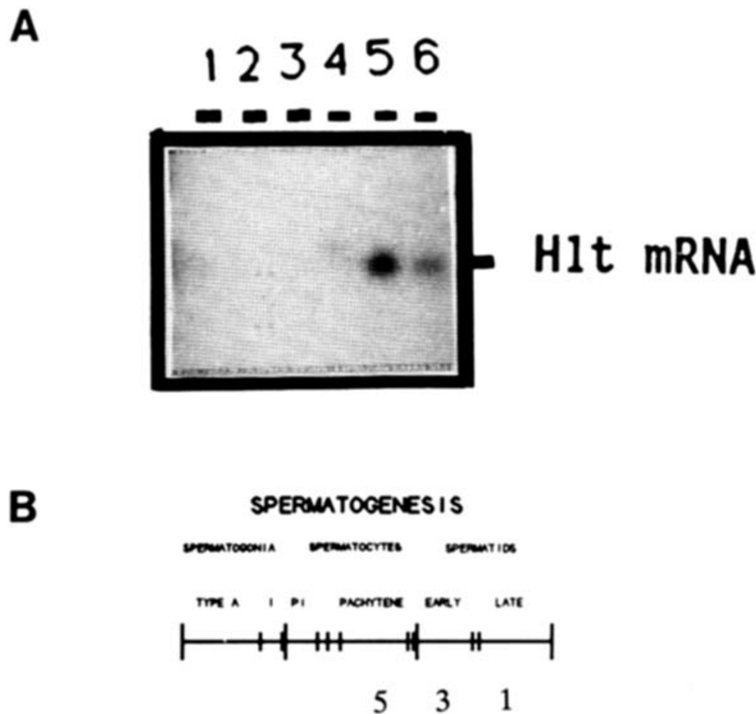

Fig. 4. Tissue-specific expression of the rat histone H1t gene determined by Northern blot analysis. Total cellular RNA samples were isolated from enriched populations of rat testis germinal cells and unfractionated rat testis. **A:** RNA samples from various rat testis cell types were electrophoresed, blotted onto NYTRAN, and probed with the 300 bp [^{32}P]-labeled PstI-HindIII fragment (nucleotides 2474–2775, Fig. 2) from the H1t gene. RNA samples (20 μg per lane) in **lanes 1–5** are derived from elutriator fractions 1–5. The two samples most enriched in RNA from specific germinal cell types are in lane 3 (70% early spermatids) and lane 5 (80% pachytene spermatocytes) (see B for a schematic representation of spermatogenesis). **Lane 6** contains 20 μg RNA from unfractionated rat testis cells. **B:** A schematic representation of the steps in spermatogenesis in the rat [45]. The two most enriched germinal cell types produced by the centrifugal elutriation technique are pachytene spermatocytes (80%) and early spermatid (70% steps 1–8). The mid-spermatid stage of short duration is the time when histone H4 is hyperacetylated and core histones and H1 histones are replaced by spermatid transition proteins TP and TP2 [5]. Midway through the late spermatid stage, the transition proteins are replaced by the rat sperm protamine designated S1 [5].

Expression of the histone H1t gene was tested by Northern blots and S1-nuclease protection analyses. These studies revealed that the H1t mRNA was present only in rat testis. Examination of rat liver and testis (Fig. 3), as well as ovary and several other tissues (data not shown), demonstrated that a measurable steady state level of histone H1t mRNA was found exclusively in testis. Further analysis indicated that a fraction of testis cells enriched in pachytene spermatocytes may contribute most of the H1t mRNA detected in testis (Fig. 4A). These data certainly do not rule out post meiotic transcription of the histone H1t gene [34], but they indicate that if histone H1t mRNA or histone H4t mRNA [18] are present in early spermatids, that they are present at very low levels. In addition, histone synthesis has been detected in premeiotic primary spermatocytes but not in postmeiotic early spermatids [20]. Therefore our

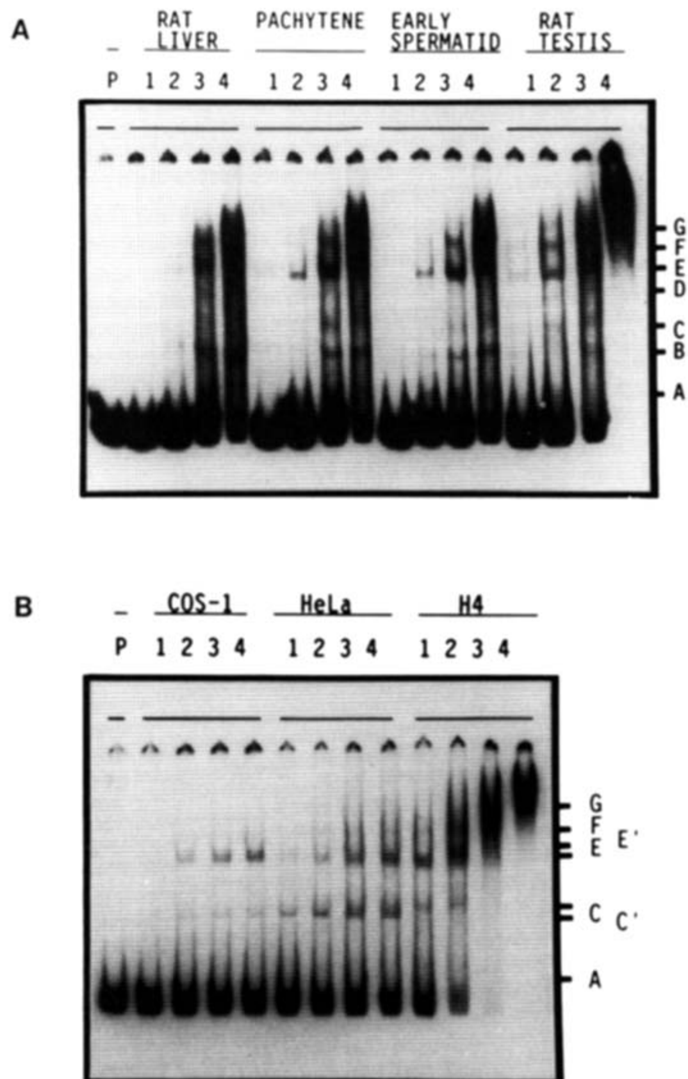

Fig. 5. Electrophoretic mobility shift assay using crude nuclear extracts and a PstI-Tth111I fragment (nucleotides 2243–2453, Fig. 2) from the proximal histone H1t promoter. The fragment was filled, dephosphorylated with calf alkaline phosphatase, and [^{32}P] end-labeled with T4 polynucleotide kinase and [γ^{32}P]-ATP. **A:** In this figure the gel retardation patterns shown were produced by binding of nuclear proteins from rat liver, from germinal cells enriched in pachytene spermatocytes (80%), germinal cells enriched in early spermatids (70%), or unfractionated rat testis cells as indicated. The first lane designated P shows the migration of free probe. The numbers **1, 2, 3,** and **4** represent progressively more crude nuclear extract in the binding reactions. For example, the amounts of protein for rat testis are 1.5, 3.0, 8.0, and 16 μg, respectively. The bands produced are designated A through G for highest mobility to lowest mobility, respectively. Note that the liver band D is absent in pachytene spermatocytes and early spermatids. **B:** In this figure the patterns shown were produced by binding of nuclear proteins from rat hepatoma H4 cells, HeLa cells, and COS-1 cells. Note that bands B and D seen in rat testis and rat liver patterns in Figure 6A are absent and two new bands C′ and E′ are present in HeLa and COS-1 cells.

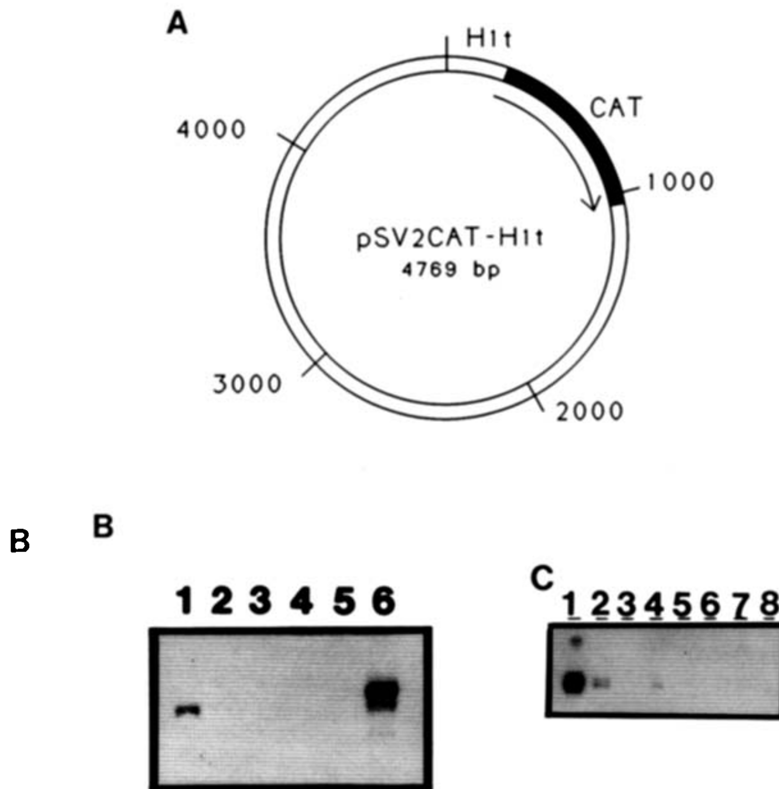

Fig. 6. Transient expression analysis using the H1t promoted CAT expression vector pSV2CAT-H1t. **A:** The H1t promoted CAT expression vector was constructed as follows. A MaeIII fragment containing the entire histone H1t gene extending from nucleotide 2211 to 3178 was Klenow filled and cloned into the SmaI site of pUC19 with the 5′-end of the H1t gene oriented toward the EcoRI side of the polylinker to produce the plasmid pJA19. The plasmid pJA19 was cut with Tth111I (nucleotide 2453 in Fig. 2) at the H1t start codon, Klenow filled, and the promoter released with EcoRI. This fragment containing the promoter was force cloned into pUC19 which had been cut with EcoRI and SmaI. The resulting plasmid was cut with EcoRI and BamHI to release the H1t promoter. The released DNA fragment containing the histone H1t promoter was Klenow filled and ligated to the plasmid vector pSV2CAT [46], which had been cut with AccI and HindIII to release the SV40 promoter and Klenow filled. Both orientations of the cloned H1t promoter were isolated in this way. Numbering starts at the beginning of the H1t promoter. The CAT gene is indicated by the filled region and is oriented as shown by the arrow. A promoterless vector pSV2CAT-Null was constructed by ligating pSV2CAT to close the vector which had been cut with AccI and HindIII to remove the promoter and which had been Klenow filled. **B:** Control Northerns with RNA from transfected cells: **1)** COS-1 transfected with 10 μg pSV2CAT-H1t; **2)** COS-1 cells transfected with 10 μg pSV2CAT-H1tR; **3–4)** blank; **5)** COS-1 cells transfected with 10 μg pSV2CAT-Null; **6)** COS-1 cells transfected with 10 μg pSV2CAT. Note that the H1t promoter is in the reverse orientation in vector pSV2CAT-H1tR used in lane 2. The SV40 promoter has been deleted in the vector pSV2CAT-Null used in lane 5. The probe used was a purified DNA fragment from the coding region of the CAT gene. **C:** Control Northerns with RNA from cells transfected with pSV2CAT-H1t: **1)** 50 μg RNA from COS-1 cells; **2)** 5 μg RNA from COS-1 cells; **3)** blank; **4)** 50 μg RNA from HeLa cells; **5)** 5 μg RNA from HeLa cells; **6)** blank; **7)** 50 μg RNA from rat H4 cells; **8)** 5 μg RNA from rat H4 cells. The probe used was the same one used in B. (Note that although there is transcription of the CAT gene in COS-1 and HeLa cells, no CAT mRNA and presumably no histone H1t promoted transcription of the CAT gene occurs in the nongerminal rat cells.)

present results are consistent with the pattern of testis-specific histone synthesis reported in previous studies.

Determination of the upstream nucleotide sequence of the rat histone H1t gene has enabled us to compare the H1t promoter with consensus elements from other histone promoters. A review of histone promoter consensus elements has been published and a comparison of the histone H1t promoter with other histone H1 promoters reveals that the testis histone H1t gene contains all of the major promoter elements found in other H1 genes [35–39]. A consensus TATA-box and an H1/CCAAT-box were found at positions 2353 and 2330, respectively, upstream from the start codon at position 2453 of the histone H1t gene. A GC-box (5′-GGGCGG-3′) is located at position 2305. This element occurs within 120 nucleotides of the transcriptional start site of some histone genes and can function in both directions. This motif, like the CCAAT-box, interacts with a transacting protein to promote maximal expression of the adjacent gene. The nuclear protein SP1 probably binds to the GC-box in this H1t gene [47].

An H1/AC-box (5′-AAACACA-3′), an H1 gene-specific element located at position 2282, is reported to be unique to and ubiquitous among H1 genes [40]. It should be mentioned that two identical copies of the H1/AC-box are located away from the proximal promoter of the histone H1t gene at nucleotide 604 at the start of the repetitive element $(CA)_{26}$ and in the distal promoter region of the histone H4t gene at nucleotide 5895. Several abbreviated copies (AAACA) of the H1/AC element and of the inverse complement (TGTTT) are located in the distal promoter regions of both the histone H1t and histone H4t genes and between the two genes. These sequences are expected to be present randomly in only about seven copies in the 6.86 kb EcoRI genomic fragment. The fact that there are 34 copies of the two sequences added together suggests that the sequence plays an important functional or structural role. The H1/AC-box has been shown to be essential for optimal expression of the histone H1 gene in S-phase [41]. However, there is no DNA replication in pachytene spermatocytes. Therefore, the function of these histone H1-specific elements and the abbreviated copies of the H1/AC element in the promoter region of the testis specific histone H1t gene are unknown.

The gene-specific motifs discussed above may not play a role in developmental or tissue specific expression of the histone H1t gene because of the global distribution of these sequences among the histone H1 genes. Other factors which influence chromatin structure or associations of the proximal or distal promoter with the nuclear matrix may play an important role in control of tissue-specific histone H1t gene expression. The erythrocyte-specific histone H5 gene contains a sequence element that is conserved in the 5′ region of the β-globin genes which is part of a binding site for erythrocyte-derived nuclear factors [42]. Such a sequence is postulated to be involved in the tissue-specific expression of the histone H5 gene. Barberis et al. [43] found a novel sequence-specific DNA binding protein that may act as a repressor of the sea urchin sperm histone H2B gene in nonspermatogenic cells. Factors in sea urchin embryos prevent CCAAT-binding protein from interacting with its sequence to increase transcription.

Examination of the rat testis specific histone H1t gene promoter revealed three additional promoter elements. The first of these is an oligo(dA) · oligo(dT) tract. An oligo(dA) · oligo(dT) tract longer than 9 bp has been reported to be involved in promoter activity [33]. The (dA) · (dT) tracts may exert their effects by causing local

perturbations in the nucleosome arrangement or structure that influence assembly of the transcriptional complex at nearby promoters. This element is known to be involved in binding to the nuclear matrix within the promoter region of several genes including histone H4 [44] and has been identified as a nuclear scaffold attachment region (SAR) in some *Drosophila* genes [32]. It should be noted that there is also an 18 bp oligo(dA) · oligo(dT) tract 1.8 kb upstream from the rat testis histone H4t gene on the same 6.86 kb genomic fragment, while a different A · T-rich sequence $(TTTA)_{11}$ occurs between the histone H1t and histone H4t genes (Fig. 1 and Table I). Secondly, there is an inverse complement of the CCAAT element (ATTGG) next to the oligo(dA) · oligo(dT) tract within the H1t promoter. The functions of the dA · dT tract and inverse complement of the CCAAT element are unknown, but they are potential candidates for regulating expression of the histone H1t gene. In this regard, preliminary in vitro DNaseI footprint analyses after binding of nuclear proteins from unfractionated rat testis or rat liver reveal partial protection within the region of the oligo(dA) · oligo(dT) tract. The third element is an 8 bp sequence (CTAGGGAT) between the GC-box (GGGCGG) and the H1/CCAAT-box within the rat histone H1t promoter. This element is part of an essential testis-specific sequence found within the promoter of the *Drosphilia* β2-tubulin gene which was shown to be indispensable for the proper testis specific expression of the gene during development in transgenic animals [31]. Therefore, it is a candidate for mediating tissue-specific transcription of the testis-specific histone H1t gene.

Data gathered thus far concerning the expression of the histone H1t gene support the hypothesis that the histone H1t gene is regulated primarily at the level of transcription. Recent data obtained in our laboratory from nuclear runoff transcription assays confirm that the testis histone H1t gene is transcribed in rat testis but not rat liver (manuscript submitted). Therefore, a better understanding of the transcriptional regulation of this gene requires an examination of protein-DNA interactions within the promoter region of the gene and with potential enhancer elements which may reside outside of the immediate promoter region. Therefore, we have initiated an investigation of the binding of nuclear proteins to the promoter region of the histone H1t gene. In order to do this, we have examined the ability of the promoter region of the histone H1t gene to bind to nuclear proteins in an electrophoretic mobility shift assay. Figure 5A,B presents results of assays with nuclear proteins from rat liver, rat testis, testis cells enriched in pachytene spermatocytes, and testis cells enriched in early spermatids, as well as nuclear proteins derived from rat hepatoma H4 cells, HeLa cells, and COS-1 cells which were in exponential growth.

The largest number of electrophoretic bands in the electrophoretic mobility shift assay is seen when the nuclear extracts are from unfractionated rat testis cells, a cell fraction enriched in pachytene spermatocytes, and a cell fraction enriched in early spermatids. These bands are designated A–G for rat testis. Band D is absent when nuclear proteins were derived from the cell fractions enriched in pachytene spermatocytes and early spermatids, but it is abundant in liver. Scanning of the autoradiograms revealed that the relative abundance of each band produced by nuclear proteins from the cell fraction enriched in pachytene spermatocytes is different from the relative abundance produced by nuclear proteins from the cell fraction enriched in early spermatids. We expect to find differences between pachytene spermatocytes and early spermatids, especially if there is a histone H1t transcriptional activator which is specific to pachytene spermatocytes, or if there is a histone H1t transcriptional repressor in

early spermatids such as the displacement protein found in nongerminal sea urchin tissues [43]. However, we could not detect the complete absence of a band or presence of a new band in the cellular fraction enriched in early spermatids compared to the cell fraction enriched in pachytene spermatocytes.

The mobility shift pattern produced with nuclear proteins from rat hepatoma H4 cells is much more similar to the testis pattern than to the liver pattern. There are only two major bands produced by nuclear proteins from HeLa cells and COS-1 cells. These bands obviously reflect the presence of nuclear proteins in the human and monkey somatic cells which bind to the testis specific histone H1t promoter, but differences in the bands may reflect species-specific or cell type differences. The binding of HeLa and COS-1 nuclear proteins to the histone H1t promoter may be functionally significant, since we have found a histone H1t promoter CAT gene to have a low level of transcriptional activity in transfected HeLa cells and COS-1 cells.

In order to produce expression vectors to test the ability of the histone H1t promoter to promote transcription, an H1t promoted-CAT gene was constructed. Control expression vectors with a reversed H1t promoter or no promoter at all were also constructed. Rat hepatoma H4 cells, HeLa cells, and COS-1 cells were transfected successfully with this H1t promoted-CAT expression vector as verified by analysis of Hirt extracts. Although we have not examined transcription directly in this study, our results suggest very low levels of transcription or repression of transcription of the rat histone H1t gene in nongerminal rat cells. Even though CAT mRNA was detected in HeLa and COS-1 cells, the levels were very low. The level of CAT mRNA in transfected COS-1 cells was much lower with the H1t promoted CAT construct pSV2CAT-H1t than with pSV2CAT. Taken together these results indicate that the region of the testis-specific histone H1t promoter used in these studies is functional but appears to promote transcription to a very low degree in transfected heterologous nongerminal cells. All of these results are consistent with repression of transcription in nongerminal cells especially nongerminal rat cells. Experiments are in progress to compare the level of CAT expression using larger histone H1t promoter sequences to the level of CAT expression using the relatively short H1t promoter sequence in this study. It is possible that DNA sequences further upstream or downstream of the histone H1t gene modify expression.

ACKNOWLEDGMENTS

We wish to acknowledge the assistance of Kenneth Wright in screening the rat genomic library and the valuable advice of Elizabeth Markose in helping establish the gel mobility shift assay. Data analyses for this study were conducted in part by using the BIONET computer resource of IntelliGenetics, Inc. (Mountain View, CA) supported by NIH grant 1 U41 RR-01685-03. This research was supported by the Medical Research Service of the Veterans Administration.

The nucleotide sequence described in this paper has been submitted to GenBank Data Library under the accession number M28409.

REFERENCES

1. Kornberg RD: Annu Rev Biochem 46:931, 1977.
2. Cole RD: Anal Biochem 136:24, 1984.

3. Cole RD: Int J Peptide Protein Res 30:433, 1987.
4. Maxson R, Cohn R, Kedes L, Mohun R: Annu Rev Genet 17:239, 1983.
5. Grimes SR: Comp Biochem Physiol 83B:495, 1986.
6. Meistrich ML, Bucci LR, Trostle-Weige P, Brock WA: Dev Biol 112:230, 1985.
7. Coles LS, Robins AJ, Madley LK, Wells JR: J Biol Chem 262:9656, 1978.
8. Sheridan WF, Stern H: Exp Cell Res 45:323, 1967.
9. Subirana JA: Exp Cell Res 63:253, 1970.
10. Hnilica LS, Johnson AW: Exp Cell Res 63:261, 1970.
11. Kaye JS, McMaster-Kaye R: Chromosom 46:397, 1974.
12. Branson RE, Grimes SR, Yonuschot F, Irvin JL: Arch Biochem Biophys 168:403, 1975.
13. Seyedin SM, Kistler WS: J Biol Chem 255:5949, 1980.
14. Shires A, Carpenter MP, Chalkley R: Proc Natl Acad Sci 72:2714, 1975.
15. Seyedin SM, Cole RD, Kistler WS: Exp Cell Res 139:399, 1981.
16. Bucci LR, Brock WA, Meistrich ML: Exp Cell Res 140:111, 1982.
17. Cole KD, Kandala JC, Kistler WS: J Biol Chem 261:7178, 1986.
18. Grimes S, Weisz-Carrington P, Daum H, Smith J, Green L, Wright K, Stein G, Stein J: Exp Cell Res 173:534, 1987.
19. Wolfe SA, Anderson JV, Grimes SR, Stein GS, Stein JA: Biochim Biophys Acta 1007:140, 1989.
20. Meistrich ML, Longin J, Brock WA, Grimes SR, Mace ML: Biol Reprod 25:1065, 1981.
21. Gorman C: (DM Glover, ed), "DNA Cloning: A Practical Approach," Vol 2. Washington, DC, IRL Press, 1985, p 143.
22. Hirt BJ: Mol Biol 26:365, 1967.
23. Struhl K: Gene 26:231, 1983.
24. Birnboim HC, Doly J: Nucleic Acids Res 7:1513, 1979.
25. Birnboim HC: Meth Enzymol 100:243, 1983.
26. Henikoff S: Gene 28:351, 1984.
27. Kraft R, Tardiff J, Krauter KS, Leinwand LA: Bio Techniques 6:544, 1988.
28. Dignam JD, Lebovitz RM, Roeder RG: Nucleic Acids Res 11:1475, 1983.
29. Shipiro DJ, Sharp PA, Wahli WW, Keller MJ: DNA 7:47, 1988.
30. Carthew RW, Chodosh LA, Sharp PA: Cell 43:439, 1985.
31. Michiels F, Gasch A, Kaltschmidt B, Renkawitz-Pohl R: EMBO J 8:1559, 1989.
32. Gasser SM, Laemmli UK: Cell 46:521, 1986.
33. Winter E, Varshavsky A: EMBO J 8:1867, 1989.
34. Moss SB, Challoner PB, Groudine M: Dev Biol 133:83, 1989.
35. Mezquita J, Connor W, Winkfein RJ, Dixon GH: J Mol Evol 21:209, 1985.
36. Van Wijnen AJ, Wright KL, Massung RF, Gerretsen M, Stein JL, Stein GS: Nucleic Acids Res 16:571, 1988.
37. Wells DE: Nucleic Acids Res 14:r119, 1986.
38. Yang Y-S, Brown DT, Wellman SE, Sittman DB: J Biol Chem 262:17118, 1987.
39. Gallinari P, La Bella F, Heintz N: Mol Cell Biol 9:1566, 1989.
40. Coles LS, Wells JRE: Nucleic Acids Res 13:585, 1985.
41. Dalton S, Wells JRE: EMBO J 7:49, 1988.
42. Wells JRE, Coles LS, Robins AJ: (LS Hnilica, ed) "Histones and Other Basic Nuclear Proteins." Boca Raton: CRC Press, Inc, 1989, p 253.
43. Barberis A, Superti-Furga G, Busslinger M: Cell 50:347, 1987.
44. Pauli U, Chiu J-F, Ditullio P, Kroeger P, Shalhoub V, Rowe T, Stein G, Stein J: J Cell Physiol 139:320, 1989.
45. Clermont NY, Harvey SC: Endrocrinology 76:80, 1965.
46. Gorman CM, Padmanabhan R, Howard BH: Mol Cell Biol 2:1044, 1982.
47. Dynan WS, Tjian R: Nature 316:774, 1985.