© 1994 Oxford University Press

Nucleic Acids Research, 1994, Vol. 22, No. 8 1421–1428

Isolation and characterization of two replication-dependent mouse H1 histone genes

Yonghe Dong, Allen M.Sirotkin, Yih-Sheng Yang[1], David T.Brown[1], Donald B.Sittman[1] and Arthur I.Skoultchi*

Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461 and [1]Department of Biochemistry, The University of Mississippi Medical Center, Jackson, MS 39216, USA

Received January 3, 1994; Revised and Accepted March 9, 1994 GenBank accession nos L26163 and L26164

ABSTRACT

Mice contain at least seven nonallelic forms of the H1 histones, including the somatic variants H1a–e and less closely related variants H1° and H1t. The mouse H1° and H1c (H1var.1) genes were isolated and characterized previously. We have now isolated, sequenced and studied the expression properties of two additional mouse H1 genes, termed H1var.2 and H1var.3. Extensive amino acid and nucleotide sequence comparisons were made between the two genes and other mammalian H1 histone genes. A high degree of nucleotide sequence identity was seen between the H1var.2, rat H1d and human H1b genes, even well beyond the coding region, indicating that these genes are likely homologues. Unlike the previously characterized mouse H1var.1 gene which produces both nonpolyadenylated and polyadenylated mRNAs, the H1var.2 and H1var.3 genes produce only typical, replication dependent, nonpolyadenylated mRNAs.

INTRODUCTION

The histones are a family of proteins that participate in organizing the structure of eucaryotic chromatin. There are five major classes of histones, H1, and the core histones H2a, H2b, H3, and H4 (1). The core histones are responsible for the formation of the nucleosomal core particle (2). H1 histones bind to the DNA between core particles and participate in the formation of higher order chromatin structures (3). There is now considerable evidence that chromatin structure is important in the regulation of transcription by controlling the accessibility of DNA to both general, and gene-specific transcription factors. Several recent studies indicate that H1 histones play a critical role in organizing these structures (4–6). Recent studies also indicate that certain antirepressors and antirepressor activities of some transcription factors interact specifically with H1s (7–9).

A characteristic of histones that is often not fully appreciated is their considerable diversity. In mammals each class of histone with the exception of H4 is represented by nonallelic primary sequence variants or subtypes which display distinct patterns of expression during development and differentiation (3, 10–12). Whereas the fundamental structural role of histones in chromatin is understood, the functional role, if any, of different histone variants in generating or maintaining different chromatin conformations within the cell is not understood. The H1 class of histones displays the most complex pattern of subtypes including differentiation-specific and tissue-specific members. At least seven different primary sequence variants have been detected in mice (10). This large number of H1 subtypes is common in higher eukaryotes. The large number and different patterns of expression among the H1 subtypes suggests that these H1 variants are in large part responsible for the wide variations in chromatin condensation that exists within the genome and in different cell types. However proof of a functional difference for the H1 variants is not available. Most approaches to testing functional significance would be greatly facilitated by the isolation of clones of the different H1 variant genes.

We have previously reported on the isolation of two mouse H1 genes encoding the H1° and H1c (H1var.1) subtypes (13–15). The H1° gene encodes a replication-independent polyadenylated mRNA, whereas the H1c gene encodes both a nonpolyadenylated, replication-dependent mRNA and a polyadenylated, replication-independent mRNA (15). We report here the isolation, sequence and expression properties of two additional mouse H1 variant genes. Both genes are expressed in a replication-dependent manner. One of the genes shows extensive homology with the rat H1d gene (16), even well beyond their transcribed regions. This gene is likely the mouse homologue of the rat H1d and one of our laboratories (DBS) has shown recently that it codes for the differently named mouse H1e variant (17). The mouse H1e gene is also linked to an H2b pseudogene. The variant type of the second mouse H1 gene could not be deduced from sequence analysis or by its expression characteristics.

*To whom correspondence should be addressed

MATERIALS AND METHODS

Isolation and sequencing of mouse histone genes

A partial MboI BALB/c mouse genomic library in lambda Charon 35 (Fred W.Alt, Harvard Medical School) was screened with the mouse cDNA clone CMH1-1.8. CMH1-1.8 was isolated by screening a lambda gt10 cDNA library from MEL cells with a human H1 histone gene probe (18). It contains a 0.63 kilobase (kb) insert with homology to H1 genes as determined by partial sequence analysis. The 0.63kb cDNA insert was purified by electrophoresis and labeled by nick translation. The labeled probe was hybridized to nitrocellulose filter lifts from about 2×10^5 plaques. Hybridization was carried out in 50% formamide/5×SSC at 42°C. Positive plaques were subjected to two further rounds of purification. Ten positive plaques were shown to fall into four groups based on their restriction enzyme digestion patterns and by Southern blot analysis with H1 histone coding region probes. One member from each of two groups, MGH1.10 (designated H1var.2) and MGH1.5 (designated H1var.3) were further characterized.

A 2.4kb AvaI fragment and an overlapping 0.6kb RsaI fragment of H1var.2 phage DNA that hybridized with the cDNA CMH1-1.8 probe were isolated and subcloned in both orientations in pBluescript SKm13+ plasmid vector (Stratagene). Similarly, two PstI fragments of 1.8kb and 0.5kb, and an overlapping 1.1kb SacI fragment from H1var.3 phage DNA were isolated and subcloned in both orientations in the same vector. A nested set of deleted inserts was generated by exonuclease III treatment for each subclone following the Erase a Base procedure (Promega). Subclones with inserts differing by about 200 nucleotides in size were detected by agarose gel electrophoresis and sequenced in both directions by the dideoxy chain termination method with Sequenase (United State Biochemical) according to the manufacturer's instructions.

The other genomic clone, MH175 which also contains the H1var.2 gene, was isolated independently by screening a partial EcoRI mouse spleen DNA library in lambda Charon 4a (R.Perlmutter and L.Hood, California Institute of Technology, Pasadena CA) with a universal H1 histone oligonucleotide probe as described previously (13).

Clone MH175 contains a 17.6 kb DNA insert. Restriction mapping localized the H1 gene to a 3.8 kb SphI–BglII fragment. This fragment was made blunt-ended with T4 DNA polymerase and subcloned into the M13mp18 vector (19) at the SmaI site. Both orientations of the 3.8 kb fragment in the M13mp18 vector were obtained as determined by the c-test (20). A nested set of deletions of clones in either orientation were made by progressive deletion with exonuclease III (21). Overlapping subclones, in both orientations, spanning the entire region containing the H1 gene were sequenced primarily by the dideoxy chain termination method. Subclones containing regions of high GC content were sequenced using reverse transcriptase. The conditions and procedures used for sequencing with reverse transcriptase were as described (22–23) with slight modification; the reactions were at higher temperatures, 42°C or 45°C, in the absence of salt in the reaction mixture.

Cell culture and RNA isolation

Clone DS19 MEL cells (24) were grown in Dulbecco modified Eagle's (DME) medium supplemented with 10% (vol./vol.) fetal bovine serum. Erythroid differentiation was induced by treatment with 5mM HMBA as described previously (14). Hydroxyurea treatment of MEL cells and serum stimulation of quiescent mouse 3T6 cells have been described previously (15). Total RNA from mouse tissues was made by the lithium chloride precipitation method (25). Polyadenylated RNA was prepared by oligo(dT)-cellulose chromatography (26) from cytoplasmic RNA made from either MEL cells or tissues.

RNase protection assays

10μg of RNA were hybridized with 10^6 dpm of uniformly labeled antisense RNA probe at 65°C for 16hrs in 10μl of 80% formamide, 0.4M NaCl, 1mM EDTA, 0.03M sodium PIPES, pH4.5. After hybridization samples were diluted into 300μl of 0.3M NaCl, 10mM Tris–HCl pH7.5, 5mM EDTA and treated with 40μg/ml RNase A and 2u/ml RNase T1 at 30°C for 30min. Samples were then incubated with 100μg proteinase K at 37°C for 15min, extracted with phenol and chloroform, precipitated with ethanol and electrophoresed on a 6% denaturing polyacrylamide gel. Assays for H1var.2 mRNA were carried out with probe a (Figure 3A), corresponding to the 3′ portion of the coding region or with probe b (Figure 3A), a genomic DNA fragment spanning the 3′ end of the gene. H1var.2 probe a was transcribed with T7 RNA polymerase after linearizing with PvuII the cDNA clone CMH1-1.8 in pBluescript SKM13+. H1var.2 probe b was similarly prepared from a deletion clone of the H1var.2 gene designated pSK MGH1.10A10(5)3. H1var.3 probes a and b (Figure 3A) were prepared with T3 RNA polymerase from the 0.5kb Pst I fragment cloned in pBluescript SkM13 + after linearization with EspI or HindIII respectively.

RESULTS

Isolation of mouse H1 histone genes

The two laboratories involved in this study have used different strategies to isolate mouse H1 histone genes. One laboratory (D.B.S.) screened a mouse genomic library with a universal H1 oligonucleotide that was intended to be unbiased for any of the H1 genes except H1°. This strategy led to the isolation of the genomic clone MH143, which was described previously (13), and MH175 which is described here. Both of these cloned genes were chosen for further study because they appeared to be linked to core histone genes.

The other laboratory (A.I.S.) screened a cDNA library, made from differentiating murine erythroleukemia cells, with a human H1 gene probe. The objective was to isolate H1 histone cDNAs appearing during cell differentiation that might be encoded by polyadenylated mRNAs. Indeed the first cDNA clones isolated in this way encoded the polyadenylated H1° mRNA and a polyadenylated form of the H1var.1 (H1c) mRNA (14,15). The H1var.1 gene is identical to the MH143 gene (13) and can produce both poly A^+ and poly A^- mRNAs (15). A third cDNA clone isolated in this way and designated cMH1-1.8 was sequenced and found by homology to code for an H1 histone protein. The cDNA insert from cMH1-1.8 was used as a probe to screen a mouse genomic library. Ten genomic clones were purified and found to fall into four distinct groups based on their restriction enzyme digestion patterns (data not shown). One group was shown by hybridization with H1var.1 specific probes to contain the H1var.1 (MH143) gene (15). A second group was identified to contain the gene from which the cMH1-1.8 cDNA clone was derived. The identification of this gene, designated H1var.2, was confirmed by sequencing and shown to be identical

to clone MH175. The third group contained an H1 gene designated H1var.3, which is also described below. The fourth group was least homologous to H1 probes by Southern blot analysis and these clones have not been further analyzed. In summary using two different isolation procedures four mouse H1 genes have been isolated: H1°; H1var.1 (MH143); H1var.2 (MH175); and H1var.3. We will use the 'H1var. #' nomenclature to refer to these genes in this manuscript.

Structure of the mouse H1var.2 and H1var.3 genes

The mouse H1var.2 and H1var.3 genes were isolated from mouse genomic DNA clones in lambda phage. Both of these genomic clones also contain a core histone gene. An H2b pseudogene is located about 720bp downstream from the end of the H1var.2 gene and is oriented in the same direction. This region was sequenced and found to be 63% identical to an expressed H2b gene. However, the putative coding region of this gene contains three mutations resulting in in-frame termination codons, multiple deletions and mutations in the putative translation initiation and termination codons. The pseudogene also contains mutated promoter elements and a mutated 3′ stem-loop sequence suggesting it did not arise through a reverse transcription intermediate. The genomic clone containing H1var.3 also contains an H3 gene as determined by Southern blot hybridization but its exact location and sequence has not yet been determined. Both the H1var.2 and H1var.3 genes appear to be single copy genes, as determined by genomic Southern blot analysis with gene specific probes (data not presented).

The complete nucleotide sequences of the H1var.2 and H1var.3 genes and their immediate flanking regions have been submitted to the GenBank™/EMBL Data Bank with accession numbers L26163 for H1var.2 (MH175) and L26164 for H1var.3. Both genes contain long open reading frames which code for proteins of 218 and 212 amino acids respectively (not counting the amino terminal methionine which is removed in H1s). Both proteins have extensive homology with known mammalian H1 amino acid sequences (see below). Typical of replication dependent histone genes the coding regions of these genes are not interrupted by introns. The transcription initiation sites were determined by S1 nuclease mapping and primer extension studies (data not presented) and are 25 and 30 nucleotides, respectively, downstream from consensus TATA boxes (27). The 5′ promoter regions of these genes each contain a consensus CCAAT box (28) as well as several potential binding sites for the SP1 transcription factor (29–30). An H1 consensus sequence (AAACACAA), which is reported to be involved in the expression and cell cycle regulation of H1 genes (31) is present at the appropriate position in the H1var.2 gene, as it is in the H1var.1 gene (15). A similar, but not identical element (AAACACGAA), is present in the H1var.3 gene. Interestingly, it is considerably farther upstream, > 60bp, than in the other H1 genes. Whether or not this is a functional H1 specific expression element is unknown. At the 3′-ends of the H1var.2 and H1var.3 genes there are highly conserved, dyad symmetry elements and consensus downstream sequence elements, essential for 3′ end processing of replication-dependent histone gene transcripts (32).

Sequence comparison of H1 histone genes

Figure 1 shows an amino acid sequence comparison of the three mouse H1 genes which we have characterized, to the human H1b [H1.4] (33,34), H1d [H1.2] (34,35), H1c [H1.3] (34,35), H1a (35), H1.1 (34) and rat H1d (16) sequences. This sequence alignment strongly suggests that the mouse H1var.2 gene is the homologue of the rat H1d and the human H1b genes. Their sequences are all 94%–96% identical to each other. Within the globular domain (36,37), the only difference among the three proteins occurs at position 121 where there is a conservative arginine to lysine change. Most of the changes between these three proteins involve conservative lysine–arginine changes or substitutions of alanine by other residues. This high degree of

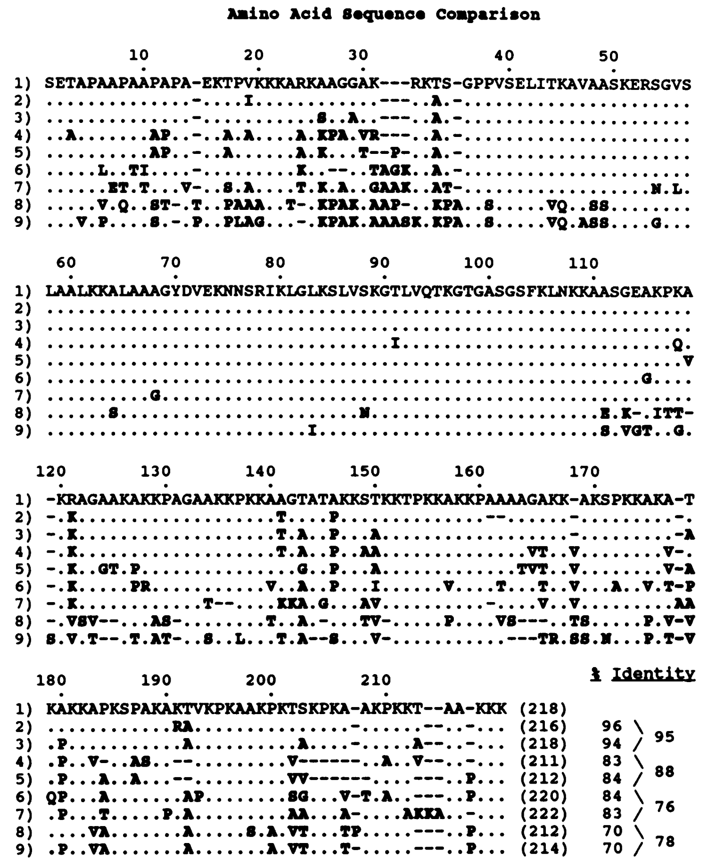

Figure 1. Amino acid sequence comparison of mouse, human, and rat H1s. All comparisons are made to the derived amino acid sequence of the mouse H1var.2 histone (lane 1). The sequences compared are: 2, the rat H1d (16); 3, the human H1b or H1.4 (33,34); 4, the mouse H1var.1 (15); 5, the human H1d or H1.2 (34,35); 6, the human H1c or H1.3 (34,35); 7, the human H1a (35); 8, the mouse H1var.3; and 9, the human H1.1 (34). Amino acid sequences are presented in the single letter code. The optimal alignments were determined manually. For the alignments a dot indicates an identity with the mouse H1var.2 sequence. A dash indicates a gap or deletion in a position where other histones have an amino acid in this position. The total number of amino acids for each protein is given in parentheses at the end of each sequence. The percent identity of each sequence as compared to the mouse H1var.2 is also given at the end of each sequence. The percent identity between other variants is presented adjacent to the bracketed variants being compared. Recently the sequence of the rat H1d gene was extended (46) and revealed the presence of a H2b pseudogene in a similar position and orientation as the mouse H2b pseudogene is to the mouse H1var.2 gene. In addition, the new rat H1d sequence contains a few differences from the previous rat H1d sequence (16) used in the comparison shown here. These differences slightly increase the homology between the rat H1d and mouse H1var.2 genes. Most importantly, the new rat H1d sequence indicates the presence of two alanine residues at positions 161 and 162 (line 2) that were absent in the previous rat H1d sequence.

A

```
     -263
      *
1)   AAATATTGCAAGTTTTTGGCAGAAGGCTTAATGCTTT-GCTCACTTTTTGAGATGTGCATGAAGCCCGAGACTTC-GGGAACATTATTTTTTAAAAATCGGG
2)   /..............C............G......-..A..T..........A.........T.T........G....T........GAG.....CT...
3)   ...................C.......TC.....CCC..T.TAA........C...G.......T...GG...G....T......C.GAG.....C....
4)                                                                                                ...C...-A

                                                           "H1-Consensus"
1)   CAATTCGATGTAGGGAATTTTGACATTTTTGGCTTTTTTTGAGGTGTAACAAACACAACTCCGCATCTGAGAGGACACT-CCGC-GGCT----GCCAGCGAGGCGG
2)   ......T....G..............................................A...........-....-....----.............
3)   ..G....G.....AC......T.T.....................................G.G...C...........-.T..-....----.............
4)   AG...T.G..C.A.AC.....A.T.....CTC......G...A...............G.TG.GCG.A.-CC..CTG..-.A.A-.CT.----.....GC....-C
5)                              C.C....G.......G.............TG.G.G.A.C.CA.CGG..-.A.A-.C..GCCA.....GCG....A
6)                                          ..-C.A......G.AGCT__60bp__....-.GGGG..TTA..G.-C----...-.G.....

                                                                   1
              "CCAAT"              "TATA"                          *
1)   GCTGGA-CAGCGCACCAATCACGGCGCAC-GTCCGCCC--TATATAAACGGGC-GGGCGCAG-CGCCGCGGCCCGAGTCCTGGCCAGTCCCTCTGCTTCCGGCTCG
2)   ......-...................G.--.......--.............-........-...G.............C.......G.................
3)   ......-...................-.......--.............-........-.........T..-....C.......G.............T.GAA
4)   ......-----..........GC....G.CTG.T..T.TA......C.GC.C.-....C...-...-TACT.TATC.G.G------------.A..G---C.A.T
5)   -.CA..-----..........GA....G.CTG.-..T.TA......C.GC...CCT..C...G...-TGCTT.ATC.G.G------------..T-.G---C.A.T
6)   ..GTT.G..TT-..........A..A..GCA-.T..T.--....A.T..TCCG--T..--..GG...TT__________54bp, TC-rich____________
7)            ................--....CA..C-....A.T.G.CTG--C.T--T......TT__________52BP____________________

                                64
                                 *       % Identity
1)   AGTT-CTCTCTCCT--CACAC--G-CTTCGCCATG
2)   ....-.........TA.....--.-....A..     93 (304/325) \
3)   .A..G.....G.-.--...G.TT.C....AA.     89 (283/326) /  87 (283/326)
4)   ....-.-..GA.T.--.CAGA--.-AC..AT.     58 (136/235) \
5)   T..A-.-.CGAGT.--TTTGATT--.-..AA.     53 (103/194) /  81 (141/174)
6)   ____________________________A..
7)   __________________________..A..
```

B

```
                                      "Histone Dyad Element"

                                           <-----    ----->                                        % Identity
1)  -ATTTT-CTTTGGCCAACTGCTTAGAAGCCCAAC--AACCC-AAAGGCTCTTTTCAGAGCCACCCACAACTCTCAGTAAAAGAG-CTGTTGCAC-ATTCAC
2)  A.....-..........................--.....-.........................G.....A........-.........-....--   [96 (90/94)] \
3)  A.AG..C.........................AC.....-........................CG...............-.........T...---   [93 (87/94)] /  [91 (85/93)]
4)  -GC.GC-..----------------CC..T...--..T..A.....................-..TCCCACTT----.....-.....-.-..-...T--
5)  -GC-GA-ACGCCT------------.CTT.T..---.....A....................-..TGA.....A........-...--.-.T-.A.TT.
6)  -ACAG----....AGG-..T...-------.T.T--.....-..C.................A...TAA...----......C...----..-.C.TGT
7)  -...CA-G..A.---..--.T..CTT---.T.GT--.....A..C.........A.........T..GCA.T....G.......-....A.T..-.CAG.T

1)  AGGGGGTGTGGTTAGGGTGGGAACAA-ACCTTAGTGGGCTGAGCTT-AGCAG---AGGAGGCTGGGGATATGTGGGTGTGC-GCAATGGTGTTAAC
2)  G.....C.................GCT...-....A..........-.-...TTA...-...........A..AC......T.-.GG...---.-.      [88 (162/184)]
3)  ......C....C.C...                                                                                  91 (101/111)
4)  -----.....A...A                                                                                    54 (59/109)

1)  TCTAGCTTAGCTATCTG-AAGGTCT-GTACAT
2)  .......-GT..CGGG.C....-..A..-T.G                                                                   85 (181/214)
```

Figure 2. DNA sequence comparison of the 5′ and 3′ flanking sequences of several mouse, human and rat H1 genes. All sequences are compared to the 5′ (A) and 3′ (B) flanking sequence of the mouse H1var.2 gene (line 1). The sequences compared are: 2, the rat H1d gene (16); 3, the human H1b or H1.4 gene (34); 4, the mouse H1var.1 gene (15); 5, the human H1d or H1.2 gene (34); 6, the mouse H1var.3 gene; and the human H1.1 gene (34). The optimal alignments were determined manually. For the alignments, a dot indicates an identity with the mouse H1var.2 gene. A dash indicates a gap or deletion in a position where other genes have a nucleotide in this position. The percent identity to the mouse H1var.2 gene, and the number of nucleotides compared to obtain this percent identity are given at the end of each line. The percent identity between other genes is presented adjacent to the bracketed gene sequences being compared.

identity in amino acid sequence is also reflected in the high degree of coding nucleotide sequence identity, 92%, in all domains, between H1var.2 and rat H1d. One of our laboratories has recently shown that the H1var.2 gene encodes the mouse H1e subtype (17). It is thought that the rat H1d protein is equivalent to the mouse H1e subtype (see Discussion).

Comparison of the nucleotide sequences in the noncoding regions of the H1var.2, rat H1d and human H1b genes further supports the likelihood that these genes are homologous (Figure 2). Beyond functionally important consensus sequences, histone genes typically exhibit few similarities between related genes of the same or different species (31). However, in the approximately 300 bp region preceding the initiation codon (the extent of 5′ end sequence available for the rat gene), the mouse H1var.2 and rat H1d genes are 93% identical and the mouse H1var.2 and human H1b genes are 89% identical.

This homology of flanking, transcribed and nontranscribed sequences also exists in the 3′ ends of the mouse H1var.2 and rat H1d genes (Figure 2B). They are 96% identical for the first 94 bp. The percentage of identity falls off when the comparison is extended for all of the available rat H1d sequences, but it is still significantly high, 85% for 214 bp. A comparable similarity exists for the mouse H1var.2 and human H1b genes. They are 91% identical for 111 bp (Figure 2B).

The amino acid sequence comparisons (Figure 1) also reveal that the human H1.2 and mouse H1var.1 are more closely related to each other, 88% identity, than either is to the mouse H1var.2, 83% and 84% respectively. Likewise, the human H1.1 and mouse H1var.3 are more closely related, 78% identity, than either is to the mouse H1var.2, 70%. Whether these two pairs of H1 variants, the mouse H1var.1 and human H1.2, and the mouse H1var.3 and human H1.1 are homologous variants, is not as clear as for the mouse H1var.2, rat H1d and human H1b. Sequence alignment of noncoding, flanking nucleotide sequences, reveals that the mouse H1var.1 and human H1.2 genes share little similarity to the mouse H1var.2 gene beyond the consensus sequences in the 5′ end: they are 58% and 53% identical to mouse H1var.2, respectively. However, that mouse H1var.1 and human H1.2 are similar is seen in their 81% identity for 174 bp of 5′ flanking sequence (Figure 2A). Although the similarity of mouse H1var.1 and human H1.2 is quite high in their 5′ regions their similarity at the 3′ end does not extend beyond the 3′ end formation and processing signals (Figure 2B).

No significant similarity is seen between mouse H1var.3 and human H1.1, either between themselves or when compared to the other mammalian H1s in either their 5′ or 3′ flanking sequences, beyond shared consensus sequences.

Evolutionarily, the mouse H1var.1 and human H1.2 are more closely related than either is to the other H1 variants. The same is true for the mouse H1var.3 and human H1.1 variants. However, in neither case is the similarity so strong, as with the mouse H1var.2, rat H1d and human H1b, to indicate that they are likely homologous variants.

Expression characteristics of the H1var.2 and H1var.3 genes

The mammalian core histone genes are generally thought to consist of at least three categories based upon their expression characteristics, particularly in relation to DNA synthesis: the replication-dependent genes in which expression is coupled to DNA synthesis; the replication-independent genes whose expression in the cell cycle occurs independently of DNA synthesis; and histone variants which are expressed in a tissue-specific manner (11,12). Whether or not these distinguishing characteristics apply to the H1 histone variants is not entirely clear at present. In fact, we have shown previously that the mouse H1var.1 gene exhibits both replication-dependent and replication-independent expression owing to the production of two different mRNA species: a short cell cycle regulated mRNA terminating in the conserved stem-loop sequence and a much longer poly A^+ mRNA that is not cell cycle regulated (15). The latter mRNA was first identified as being produced during induction of mouse erythroleukemia cell differentiation. We showed it is also found in low amounts in a number of mouse tissues, and becomes more predominant in adult mouse tissues, especially in the kidney where it becomes the dominant H1var.1 mRNA species (15). Therefore, it was of interest to determine whether the expression of the H1var.2 and H1var.3 genes is coupled to DNA replication and whether the genes might also produce poly A^+ mRNAs.

To assay expression of the two H1 genes we utilized a sensitive RNase protection assay with uniformly labeled RNA probes that extend from the 3′ portion of the coding region into the 3′ flanking sequences beyond the stem-loop element (Figure 3A). When such probes are hybridized with RNA, mRNA terminating

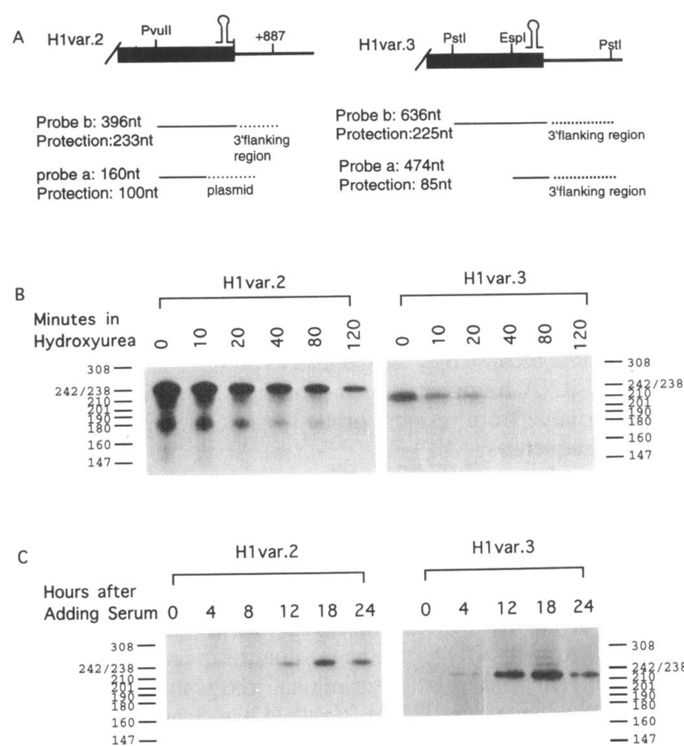

Figure 3. H1var.2 and H1var.3 transcript levels in relation to the cell cycle. **(A)** Schematic diagram of RNA probes used in RNase protection assays. The solid lines depict the regions of the probe expected to be protected by mRNAs terminating at the histone dyad symmetry element, whereas the dotted lines represent regions of the gene and plasmid DNA sequences beyond the expected mRNA termini. The lengths of the probes and of the expected fragments protected by mRNAs terminating at the dyad symmetry element are indicated. **(B)** MEL cells were treated with 5mM HMBA for 8 hrs and then exposed to 5mM hydroxyurea. Total RNA was isolated at the times indicated above the lanes. 10μg of RNA were assayed by RNase protection using H1var.2 probe b and H1var.3 probe b. **(C)** Quiescent mouse 3T6 cells were stimulated by addition of 10% serum and total RNA was isolated at the times indicated above the lanes and assayed by RNase protection as in (B).

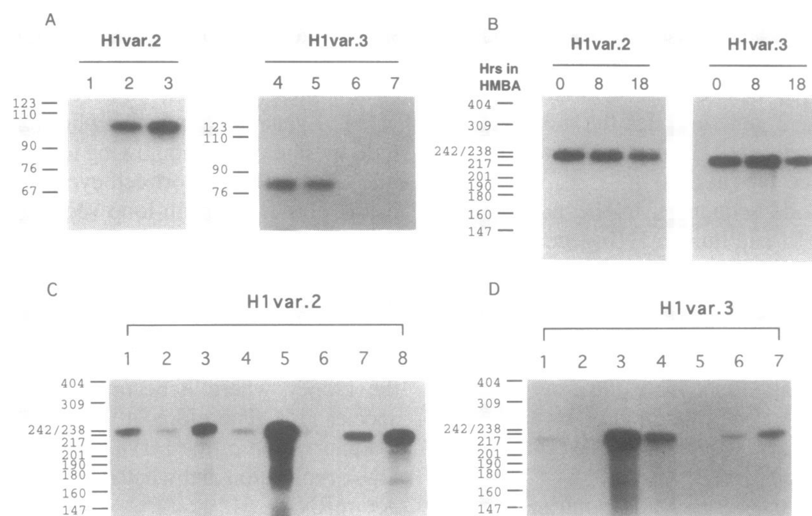

Figure 4. H1var.2 and H1var.3 transcript levels in RNA fractions from oligo(dT)-cellulose chromatography and from MEL cells and mouse tissues. (**A**) Total RNA made from uninduced MEL cells was fractionated by oligo(dT)-cellulose chromatography. RNA samples from bound poly(A)$^+$ RNA (0.1μg, lanes 1, 6; 1μg, lane 7); unbound poly(A)$^-$ RNA (10μg, lanes 2, 5), and unfractionated RNA (10μg, lanes 3, 4), were hybridized with antisense RNA probes for H1var.2 (probe a, lanes 1, 2, 3) and H1var.3 (probe a, lanes 4, 5, 6,7) shown in Figure 3A. (**B**) Total cellular RNA was prepared from MEL cells grown in the presence of 5mM HMBA for the times indicated (in hours) above each lane. RNA was hybridized with antisense RNA probes transcribed from H1var.2 (probe b) or H1var.3 (probe b) as shown in Figure 3A. (**C**) Total RNA was prepared from dissected livers (lanes 1, 2), kidneys (lanes 3, 4), thymus (lanes 5, 6) and testes (lanes 7, 8) of 2-week-old (lanes 1, 3, 5, 7) and 9-week-old (lanes 2, 4, 6, 8) BALB/c mice. 10μg of RNA were analyzed in RNase protection assays using H1var.2 probe b. (**D**) Total RNA was prepared from livers (lanes 1, 2) thymus (lanes 3, 4), kidneys (5) and lungs (lanes 6, 7) of 2-week-old (lanes 1, 3, 7) and 9-week-old (lanes 2, 4, 5, 6) BALB/c mice. 10μg of RNA were analyzed by RNase protection using H1var.3 probe b. In panels B, C, and D the region of the autoradiogram above the 404 bp marker fragment is not shown. No protected RNA fragments were visible in this region in any of the samples.

near the stem-loop element will produce a short protected RNA fragment, whereas any mRNA species terminating beyond the stem-loop element (e.g., a poly A$^+$ mRNA) will produce longer protected RNA fragments. Even if such longer mRNA species terminate beyond the sequences present in the probes they would be detectable because they should lead to production of full-length protected RNA fragments. Such protected RNA fragments are distinguishable from residual probe because the probes include plasmid sequences.

When cycling mammalian cells are treated with hydroxyurea, an inhibitor of DNA replication, the levels of replication-dependent histone mRNAs decline rapidly due primarily to posttranscriptional destablization of the mRNAs (38). Conversely, when quiescent cells are stimulated to re-enter the cell cycle by addition of growth factors, the levels of these mRNAs rise in conjunction with DNA synthesis due to a combination of transcriptional and posttranscriptional mechanisms (39). To determine whether the H1var.2 and H1var.3 genes exhibited replication-dependent expression we analyzed the levels of their mRNAs following both inhibition and stimulation of DNA replication. The results shown in Figure 3 demonstrate that expression of H1var.2 and H1var.3 mRNAs terminating just beyond the stem-loop sequence is coupled to DNA replication. The levels of both mRNAs expressed in uninduced MEL cells were reduced when the cells were treated with the DNA synthesis inhibitor hydroxyurea (Figure 3B). Furthermore, when quiescent mouse 3T6 cells were stimulated to reenter the cell cycle by addition of 10% serum, the levels of both mRNA's rose rapidly, at about the time that the cells began to replicate their DNA (Figure 3C). The same RNA samples were assayed in previous studies (15) for the levels of H3 histone mRNA and the poly A$^-$ and poly A$^+$ forms of H1var.1 mRNA. The results shown in Figure 5 are very similar to those observed for H3 mRNAs and the poly A$^-$ H1var.1 mRNA. In contrast, the level of poly A$^+$ H1var.1 mRNA does not change substantially during hydroxyurea treatment or serum stimulation (15).

As mentioned, the alternative, poly A$^+$ form of the H1var.1 mRNA is most readily detected in certain adult mouse tissues (e.g., kidney) and also during the induction of MEL cell differentiation after about 8 hours of inducer treatment. Therefore, we assayed RNA from these different types of cells to determine whether the H1var.2 and H1var.3 genes could produce a poly A$^+$ mRNA under the same circumstances. Total RNA was prepared from uninduced and induced MEL cells during the period when induction of the H1var.1 poly A$^+$ mRNA is readily detected (15). Total RNA was also prepared from several tissues of two week old and nine week old mice, including samples from adult kidneys in which the H1var.1 poly A$^+$ mRNA is most abundant (15). These RNA samples were assayed by RNase protection using probes from both the H1var.2 and H1var.3 genes that would be capable of detecting longer polyadenylated forms of the mRNAs. The results shown in Figures 4 B,C,D indicate that the probes derived from both genes detected only mRNA terminating at the stem-loop element in several mouse tissues and in induced MEL cells. No evidence for longer transcripts terminating beyond the stem-loop element was seen. Furthermore, consistent with the replication-dependent expression of both genes, we found that the levels of both mRNAs were higher in most tissues of newborn mice which are undergoing growth through cell division than in the tissues of adult mice in which cell division has slowed considerably. Finally we sought to determine whether any polyadenylated transcripts

of the two genes could be detected by fractionation of total RNA from MEL cells on oligo-dT cellulose. mRNAs from the two genes were only detected in the unbound, poly A^- fraction and again there was no evidence for the existence of mRNA species terminating beyond the histone dyad symmetry element (Figure 4A). We conclude that in the several conditions we have assayed, the two H1 genes produce only typical cell cycle regulated poly A^- histone mRNAs that are processed at the 3′ terminal palindrome.

DISCUSSION

Previously, our labs independently isolated the mouse H1var.1 (MH143) gene. This H1 gene was originally identified as being fully replication-dependent (13). However, it was later learned that this gene could also make a mRNA which was polyadenylated and therefore it is capable of both replication-dependent and replication-independent expression (15). We have subsequently shown using transformation studies that this gene codes for the H1c variant (17). In this paper we describe the cloning, sequencing and characterization of two additional mouse H1 variant genes. Unlike the previously described H1var.1 (H1c) gene, these genes appear to be principally expressed in a replication-dependent manner.

There is little consistency between different labs in the nomenclature and identification of H1 variants. Because of their extensive work with mammalian H1s, and the mouse H1s in particular, we have chosen to use the H1 nomenclature of Lennox and Cohen (40). Kistler's group has attempted to keep their nomenclature and identification of rat H1s consistent with the Lennox and Cohen system. Both groups base their H1 letter designation on the order of elution of the different variants from Bio Rex 70 chromatography (40). Because Bio Rex 70 chromatography is too cumbersome for routine fractionation of H1 variants each lab uses gel electrophoresis through acid and SDS gels to resolve the different H1 variants. However each lab uses different acid gel systems on which Hle and Hld show different mobilities. Because of this, the designation of H1d and H1e is reversed between the different labs. The extensive identity of the mouse H1var.2 gene sequence presented in this paper and that of the rat H1d gene, even well beyond the coding region, leads us to believe they are homologous genes. We therefore believe the mouse H1var.2 gene codes for an H1e variant by Lennox and Cohen's nomenclature system. We have recently confirmed by transformation studies that the H1var.2 gene codes for H1e (17). Amino acid sequence comparison indicates that the human H1 gene, designated H1b or H1.4 (33,34), is probably homologous to the rat H1d and the mouse H1e.

Both the H1c and H1e variants are major, abundant variants (10). It is therefore not surprising that the cDNA screening approach led to the isolation of their genes even though the H1e gene, unlike the H1c gene, does not seem to produce a polyadenylated message. cDNAs are not always produced by initiation of first strand synthesis from a poly A tail.

It has already been noted that there is a paucity of H1 genes in mammalian histone gene clusters (13) and, unless they reside in separate chromosomal locations, there may be fewer H1 genes than there are core histone genes. Both the H1var.2 and H1var.3 genes as well as the H1var.1 (H1c) and H1° genes (13,15,41) appear to be single copy genes. It is likely that, unlike the core histone genes where there may be multiple genes for the same variant, there may be fewer copies of each H1 variant gene.

Protein analysis indicates that both H1c and H1e are synthesized in dividing and nondividing cells (10). Because the H1var.1 (H1c) gene can produce either a replication-dependent or replication-independent form of histone mRNA (15), one gene is sufficient for the synthesis of H1c at all times. We found no polyadenylated mRNA produced from the H1var.2 (H1e) or H1var.3 genes and their mRNAs appear to be regulated in a replication-dependent manner. When DNA synthesis is inhibited the steady state levels of replication-dependent histone mRNA declines to 5−10% of their levels in dividing cells (42). After prolonged inhibition of DNA synthesis these basal levels of mRNA increase (42). It is possible that these low levels of mRNA in conjunction with the high stability of the H1e protein (10) is sufficient to account for the abundance of H1e in nondividing cells. It is also possible that there is a replication-independent H1e gene much like the histone H3.3 gene. However, if another H1e gene exists it must have a very different codon usage from the H1e gene described here and therefore failed to hybridize with a probe made from this gene.

Sequence analysis of the 5′ flanking sequences of the H1var.1 (H1c), H1var.2 (H1e) and H1var.3 genes reveals putative transcriptional regulatory sequences. These genes have the canonical TATA-box, CCAAT-box and SP1 binding site sequences. In addition, like the H1var.1 gene, the H1var.2 gene has a sequence at the appropriate location which matches an H1-specific sequence (31). The region surrounding this H1-specific sequence is also highly conserved between the mouse H1var.1 and H1var.2 genes. The H1var.3 gene however does not have this H1 specific sequence in the appropriate position. It has a sequence 60 base pairs upstream of the expected location which is similar (it has a one bp insertion). Whether this sequence is functional or coincidental cannot be determined by sequence analysis.

Sequence comparison of the mouse H1var.2 (H1e) and H1var.1 (H1c) genes reveals some interesting features. Unlike the comparison of the mouse H1var.2 gene to the rat H1d gene, which shows a fairly even distribution of their few differences throughout the gene and flanking region, comparison of the two mouse genes shows uneven distribution of sequence differences. The central globular domain of H1s is highly conserved (36,37). It is not surprising to find a high level of identity between these regions in different H1 variants. What is striking is that there are two stretches of reasonable length, 144 bp and 54 bp, with perfect identity. Sequence comparison of the whole central domain also shows that there are fewer silent base changes than are seen upon comparison of the amino and carboxy domains. 65% and 46% of the codons for 23 and 57 identical amino acids in the amino and carboxy termini, respectively, have silent changes compared to 12% of 102 codons in the globular domain. Both the long stretches of perfect identity and the overall high level of identity in the central domain may reflect a functional significance of the DNA or RNA sequences in this region. Proportionately long sequences of perfect identity relative to the total coding length have been observed between different replication-dependent H3 variants (43). The replication-dependent histone genes are regulated transcriptionally and posttranscriptionally (38). The 3′ stem-loop is clearly a signal recognized in regulation of posttranscriptional stability (38). Other coding or RNA sequences may be regulatory. There is a report that coding sequences in a H3 gene are necessary for proper transcriptional regulation (44) These highly conserved sequences in H1s are candidates for other regulatory signals. However, if

these sequences are significant their importance does not extend to the H1var.3 gene which shows no long stretch of identity with either the mouse H1var.1 or the H1var.2 genes.

In conclusion we have cloned and characterized two additional mouse H1 variant genes. These genes in conjunction with the mouse H1° and H1c genes and the rat and human H1t as well as other unidentified mammalian H1 genes (34,45) probably means that genes for most of the mammalian H1 variants have been obtained.

The nucleotide sequences reported in this paper have been submitted to the GenBank™/EMBL Data Bank with accession numbers L26163 for H1var.2 (MH175) and L26164 for H1var.3.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health Grant CA 16368 to A.I.S., National Science Foundation Grant DMB-8818958 and a grant from the Mississippi Affiliate of the American Heart Association to D.B.S. and by an Albert Einstein College of Medicine Core Grant (NIH 2 P30 CA13330) A.M.S. was supported by NIH Training Grant GM 07491.

REFERENCES

1. Isenberg, I. (1979) *Annu. Rev. Biochem.* **48**, 159–191.
2. McGhee, J.D., and Felsenfeld, G. (1980) *Annu. Rev. Biochem.* **49**, 1115–1156.
3. van Holde, K.E. (1988) *Chromatin. Springer-Verlag, New York.*
4. Croston, G.E., Kerrigan, L.A., Lira, L.M., Marshak, D.R., and Kadonaga, J.T. (1991) *Science* **251**, 643–649.
5. Shimamura, A., Sapp, M., Rodriguez-Campos, A., and Worcel, A. (1989) *Mol. Cell Biol.* **9**, 5573–5584.
6. Zlatnova, J. (1990) *Trends in Biol. Sci.* **15**, 273–276.
7. Laybourn, P.J., and Kadonaga, J.T. (1991) *Science* **254**, 238–245.
8. Laybourn, P.J., and Kadonaga, J.T. (1992) *Science* **257**, 1682–1685.
9. Croston, G.E., Laybourn, P.J., Paranjape, S.M., and Kadonaga, J.T. (1992) *Genes and Devel* **6**, 2270–2281.
10. Lennox, R.W., and Cohen, L.H. (1984) *Histone Genes: Structure, Organization, and Regulation.* pp. 373–395. (G. S. Stein, J. L. Stein, and W. F. Marzluff, eds.), John Wiley and Sons. New York.
11. Wu, R.S., West, M.H.P., and Bonner, W.M. (1984) *Histone Genes: Structure, Organization, and Regulation.* pp. 457–483. (G. S. Stein, J. L. Stein and W. F,. Marzluff, eds.) John Wiley and Sons. New York.
12. Zweidler, A. (1984) *Histone Genes: Structure, Organization, and Regulation.* pp. 339–371. (G. S. Stein, J. L. Stein and W. F. Marzluff, eds.). John Wiley and Sons. New York.
13. Yang, Y-S., Brown, D.T., Wellman, S.E., and Sittman, D.B. (1987) *J. Biol. Chem.* **262**, 17118–17125.
14. Cheng, G., and Skoultchi, A.I. (1989) *Mol. Cell. Biol.* **9**, 2332–2340.
15. Cheng, G., Nandi, A., Clerk, S., and Skoultchi, A.I. (1989) *Proc. Natl. Acad. Sci. USA* **86**,7002–7006.
16. Cole, K.D., Kandala, J.C., Kremer, E., and Kistler, W.S. (1990) *Gene* **89**, 265–269.
17. Brown, D.T., and Sittman, D.B. (1993) *J. Biol. Chem.* **268**, 713–718.
18. LaBella, F., Zhang, R., and Heintz, N. (1988) *J. Biol. Chem.* **263**, 2115–2118.
19. Yanish-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103–119.
20. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
21. Heinkoff, S. (1984) *Gene* **28**, 351–359.
22. Graham, A., Steven, J., McKechnie, D., and Harris, W.J. (1986) *Focus* **8**, 4–5.
23. Karanthanasis, S. (1982) *Focus* **4**, 6–7.
24. Lachman, H.M., Cheng, G., and Skoultchi, A.I. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6480–6484.
25. Bothwell, A., Yancopoulos, G.D., and Alt, F.W. (1990) *in Methods for Cloning and Analysis of Eukaryotic Genes* pp. 15–16. Jones and Bartlett, Boston.
26. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed.) pp. 7–26. Cold Spring Harbor Laboratory Press, New York.
27. Breathnach, R., and Chambon, P. (1980) *Annu. Rev. Biochem.* **50**, 349–383.
28. McKnight, S., and Tjian, R. (1986) *Cell* **46**, 795–805.
29. Kadonaga, J.T., Jones, K.A., and Tjian, R. (1986) *TIBS* **11**, 20–23.
30. Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tjian, R. (1986) *Science* **234**, 47–52.
31. Cole, L.S., and Wells, J.R.E. (1985) *Nucleic Acids Res.* **13**, 585–594.
32. Birnstiel, M.L., Busslinger, M., and Strub, K. (1985) *Cell* **41**, 349–359.
33. Ohe, Y.H., Hayashi, H., and Iwai, K. (1986) *J. Biol. Chem.* **100**, 359–368.
34. Albig, W., Kardalinou, E., Drabent, B., Zimmer, A., and Doeckne, D. (1991) *Genomics* **10**, 940–948.
35. Ohe, Y., Hayashi, H., and Iwai, K. (1989) *J. Biochem.* **106**, 844–857.
36. Allan, J., Hartman, P.G., Crane-Robinson, C., and Aviles, F.X. (1980) *Nature* **288**, 675–679.
37. Allan, J., Mitchell, T., Harborne, N., Bohm, L., and Crane-Robinson, C. (1986) *J. Mol. Biol.* **187**, 591–601.
38. Marzluff, W.F., and Pandey, N.B. (1988) *TIBS* **13**, 49–52.
39. Harris, M.E., Bohni, R., Scheiderman, M.H., Ramamurthy, L., Schumperli, D., Marzluff, W.F. (1991) *Mol. Cell. Biol.* **5**, 2416–2424.
40. Lennox, R.W., Oshima, R.G., and Cohen, L.H. (1982) *J. Biol. Chem.* **257**, 5183–5189.
41. Doenecke, D., and Tonjes, R. (1986) *J. Mol. Biol.* **187**, 461–464.
42. Sittman, D.B., Graves, R.A., and Marzluff, W.F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1849–1953.
43. Taylor, J.D., Wellman, S.E., and Marzluff, W.F. (1986) *J. Mol. Evol.* **23**, 242–249.
44. Hurt, M.M., Bowman, T.L., and Marzluff, W.F. (1991) *Mol. Cell. Biol.* **11**, 2929–2936.
45. Drabent, B., Kanadlinou, E., and Doenecke, D. (1991) *Gene* **103**, 263–268.
46. Drabent, B., Kunz, C., and Doenecke, D. (1993) *Biochim. Biophys. Acta* **1172**,193–196.