Volume 17 Number 22 1989 Nucleic Acids Research

H2A.X. a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3′ processing signals

Cecilia Mannironi, William M.Bonner* and Christopher L.Hatch

Laboratory of Molecular Pharmacology, Division of Cancer Treatment, Building 37, Room 5D19, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received August 14, 1989; Revised and Accepted October 17, 1989 EMBL accession no. X2885

ABSTRACT

A full length cDNA clone that directs the *in vitro* synthesis of human histone H2A isoprotein H2A.X has been isolated and sequenced. H2A.X contains 142 amino acid residues, 13 more than human H2A.1. The sequence of the first 120 residues of H2A.X is almost identical to that of human H2A.1. The sequence of the carboxy-terminal 22 residues of H2A.X is unrelated to any known sequence in vertebrate histone H2A; however, it contains a sequence homologous with those of several lower eukaryotes. This homology centers on the carboxy-terminal tetrapeptide which in H2A.X is SerGlnGluTyr. Homologous sequences are found in H2As of three types of yeasts, in Tetrahymena and Drosophila. Seven of the nine carboxy-terminal amino acids of H2A.X are identical with those of S. cerevisiae H2A.1. It is suggested that this H2A carboxy-terminal motif may be present in all eukaryotes. The H2A.X cDNA is 1585 bases long followed by a polyA tail. There are 73 nucleotides in the 5′ UTR, 432 in the coding region, and 1080 in the 3′ UTR. Even though H2A.X is considered a basal histone, being synthesized in G1 as well as in S-phase, and its mRNA contains polyA addition motifs and a polyA tail, its mRNA also contains the conserved stem-loop and U7 binding sequences involved in the processing and stability of replication type histone mRNAs. Two forms of H2A.X mRNA, consistent with the two sets of processing signals were found in proliferating cell cultures. One, about 1600 bases long, contains polyA; the other, about 575 bases long, lacks polyA. The short form behaves as a replication type histone mRNA, decreasing in amount when cell cultures are incubated with inhibitors of DNA synthesis, while the longer behaves as a basal type histone mRNA.

INTRODUCTION

Mammalian cells contain four histone H2A isoprotein species, called H2A.1, H2A.2, H2A.Z and H2A.X (1). In proliferating cells, the former two, called replication type, are synthesized primarily in concert with chromatin replication, while the latter two, called basal or replacement type, are synthesized throughout the cell cycle (2). In quiescent cells, all four isoprotein species are synthesized at much lower rates but in relative amounts similar to those in proliferating cells (3). H2A.X represents about 10−15% of total H2A in most mammalian cell lines and tissues, but has been reported to be a somewhat larger fraction in testes and brain (4). Evidence from analysis of tryptic peptides suggests that the H2A.X and H2A.1 sequences are similar in the amino-terminal and core regions, but that they differ considerably in the carboxy-terminal region (5).

In this study, we report the sequence of a full length cDNA clone for human H2A.X. The cDNA directs the *in vitro* synthesis of a protein that comigrates with authentic H2A.X on two-dimensional AUT-AUC polyacrylamide gels for histones (6). The H2A.X protein is composed of 142 amino acids, 13 more than H2A.1. The amino acid sequences of these two isoproteins are almost identical up to residue 120, but appear to be unrelated beyond

120. However, the H2A.X carboxy-terminal sequence is similar to ones reported in several lower eukaryotes. The H2A.X cDNA is 1585 bases long followed by a polyA tail, but also contains highly conserved sequences implicated in mRNA processing and stability (8–12). Strikingly, two sizes of H2A.X mRNA, consistent with the two sets of processing signals, are found in exponentially growing cell cultures.

MATERIALS AND METHODS

Screening and Clone Selection

Human basal histone cDNA sequences were selected from a human cDNA library constructed in an Okayama-Berg vector (7, protocols and ingredients from Pharmacia) with the polyA mRNA fraction of HeLa cells that had been incubated with the inhibitors of DNA synthesis, hydroxyurea and aphidicolin, 30 min prior to RNA isolation. Generally, replication-linked histone mRNAs lack polyA tails and are degraded when DNA synthesis is inhibited; thus this protocol provides a two step enrichment for replication-independent histone mRNAs (13–15). The cDNA library was screened with a human H2A.1 probe (see below), radioactively labeled using random primers (Pharmacia) and purified by chromatography through an Elutip-d column (Schleicher and Schuell). Twelve positive clones were selected using moderate rather than high stringency conditions for the post-hybridization washes. Positive clones were sequenced with a synthetic primer to a sequence in the Okayama-Berg vector close to the 5′ end of the cDNA. In most cases the sequence went far enough to enter an open reading frame. From peptide mapping of mouse H2As, it was known that the H2A.1 tryptic peptide gln-gly-gly-lys starting at position 6 of the amino acid sequence was changed in H2A.X (5). A clone, called 12a, was found which had a threonine in place of the glutamine in that peptide. With an internal primer to 12a, two other slightly shorter H2A.X clones with identical primed DNA sequences were isolated.

Sequencing Strategy

This putative H2A.X cDNA was excised from the Okayama-Berg construct with Bam H1; the complete cDNA and Sau 3A fragments of it were subcloned into Bluescript KSM vector (Stratagene) for sequencing. The complete sequence of both strands of this cDNA was done by the Sanger dideoxy sequencing method (16) using vector-specific T3 and T7 primers as well as additional synthetic primers complementary to specific sequences of the cDNA. The sequencing reaction used Sequenase (USB Corp.); sometimes the dGTP analogs, dITP and 7-deaza dGTP, were substituted to overcome GC compressions of the DNA sequence (17,18). In order to overcome the worst compressions, single strand DNA binding protein was added during the labeling reaction (19). Briefly, 0.5 μg of E. coli single-strand binding protein (USB Corp.) was added to the labeling reaction of the sequencing protocol. After adding the stop solution, 0.1 μg of proteinase K was added and the mixture incubated at 65°C for 20 min.

Human H2A.1 cDNA

A genomic clone for human H2A.1 had been previously isolated and partially sequenced in this laboratory (20). Sequences in this clone were identical to that of the genomic clone isolated by Zhong et al. (21, GenBank:Humhish2a), and included the region between bases 145 and 866 of that clone; the H2A coding region is from base 256 to 648.

In Vitro Transcription and Translation

In vitro transcription and translation were performed as described in Hatch and Bonner (22) using the above-mentioned constructions of the complete cDNA in Bluescript KSM.

The *in vitro* transcription and translation products of the H2A.X cDNA were analyzed on two-dimensional AUT-AUC gels for histones (6).

Preparation of RNA

Cytoplasmic total RNA was purified from exponentially growing Jurkat cell cultures (a human lymphoblastic tumor line) or from duplicate cultures that had been preincubated with 10 mM hydroxyurea for 30 min to inhibit DNA replication (2). Washed cells were suspended at 10^8 per ml in cold TES buffer (10 mM TrisHCl pH 7.5, 1 mM EDTA, 100 mM NaCl) with 10 mM VRC (23). Detergent NP40 was added to 0.25% to lyse the cells at 4°C. After nuclei were removed by centrifugation, the supernatant was made 1% in SDS, 10 mM in EDTA, 100 mM in Tris-HCl pH 8.0, and 250 mM in NaCl by adding 0.05 vol of 20× concentrated stock solutions. Predigested Proteinase K was then added to 0.2 mg/ml. The solution was incubated at 37°C for 30 min, then extracted once with phenol-chloroform. RNA was ethanol precipitated, dissolved in sterilized deionized water and stored at −70°C. PolyA selected RNA was prepared on an oligo-dT column (Collaborative Research Incorporated) (24). RNA was electrophoresed in 1.2% agarose gels using MOPS buffer with 3% formaldehyde in the gel (23). RNA was electroblotted onto Zetaprobe membranes (Bio-Rad Laboratories), which were baked under vacuum for 3 hrs at 80°C.

Hybridization of Northerns

Probes were oligolabeled (kit from Pharmacia), and purified through Elutip-d (Schleicher and Schuell) or push columns (Stratagene). Membranes were incubated for several hr at 42°C in 5 ml of prehybridization solution (50% formamide, 5× SSC, 50 mM sodium phosphate pH 6.5, 5× Denhardts mix, 0.1% SDS, and 50 μg/ml of yeast tRNA) in 50 ml screw-capped tubes mounted on a rock and roll machine set so that the interior surface was periodically bathed with solution. Labeled probe ($>10^6$ cpm/ml) was mixed with hybridization solution (50% formamide, 5× SSC, 20 mM sodium phosphate pH 6.5, 1× Denhardts mix, 0.1% SDS, 50 μg/ml of yeast tRNA, and 8% dextran sulphate). Membranes were hybridized overnight at 42°C, then washed with 0.1% SDS, 0.1× SSC for several hr at 65°C.

Mapping of 5' Terminus

A 24 base primer was synthesized (Midland Certified Reagent Company) as the reverse complement to bases 18−41 of the cDNA. The primer was end-labeled with $^{32}PO_4$ and purified through a push column (Stratagene). The primer extension protocol was as described by Kingston (23). The labeled primer was also used to sequence this region of clone 12a, by eliminating the labeling reaction (Sequenase, USB Corp.). Products were analyzed on 6% polyacrylamide sequencing gels.

RESULTS

H2A.X Protein Sequence

The human H2A.X cDNA encodes a 142 amino acid protein with a deduced molecular weight of 15012 (Figure 1). When transcribed and translated *in vitro*, the cDNA yielded a product which comigrated with H2A.X on two-dimensional AUT-AUC gels for histones (6; Figure 2). The rabbit reticulocyte cell-free translation mix contains the components for the ubiquitination of proteins, and ubiquitinated H2A.X could be seen on long exposures of these gels.

The H2A.X amino acid sequence can be divided into two regions (Figure 1). The first region includes the first 120 amino acid residues and has a 96% homology with the same

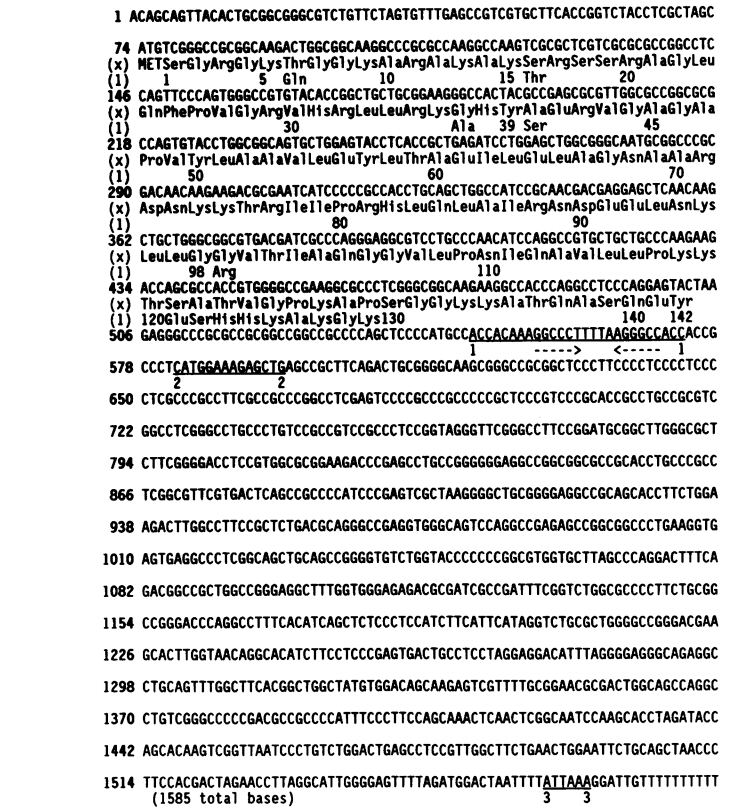

Figure 1. Sequence of human H2A.X cDNA. The nucleotide sequence is presented with the encoded H2A.X protein sequence. Where the H2A.1 (see Methods) amino acid sequence differs from that of H2A.X, the H2A.1 residue is noted beneath the H2A.X sequence. The amino acid residues are numbered at every tenth as well as near these amino acid differences. Certain nucleotide motifs in the 3′ UTR are underlined and noted with numbers (1−3). The stem-loop sequence is noted with arrows.

region in human H2A.1. There are five amino acid differences between H2A.X and H2A.1 all of which are found as substitutions in H2As of different species or in different H2A.1/2 type isoproteins of the same species (25). The threonine at position 6 used to identify the protein as a potential H2A.X, although not reported for mammalian H2A, is found in trout H2A.1 (26).

The second H2A.X region, carboxy-terminal to residue 120, is different from that in human H2A.1 in length as well as in sequence. This region in H2A.X is among the longest found in sequenced H2As, being 13 residues longer than that of H2A.1. The striking feature of the human H2A.X carboxy-terminal sequence is its homology with the carboxy-terminal sequences of several species of lower eukaryotes (Figure 3); seven of the final nine amino acid residues, H2A.X positions 134 to 142, are identical to those of S. cere. H2A.1 (27). Similar carboxy-terminal amino acid sequences are present in H2As of two other fungi, A. nidulans (28) and S. pombe (29,30) as well as in one of the Tetrahymena H2As (31). In all these sequences, there is a conserved (Ser-Gln-acidic-aliphatic) tetrapeptide motif

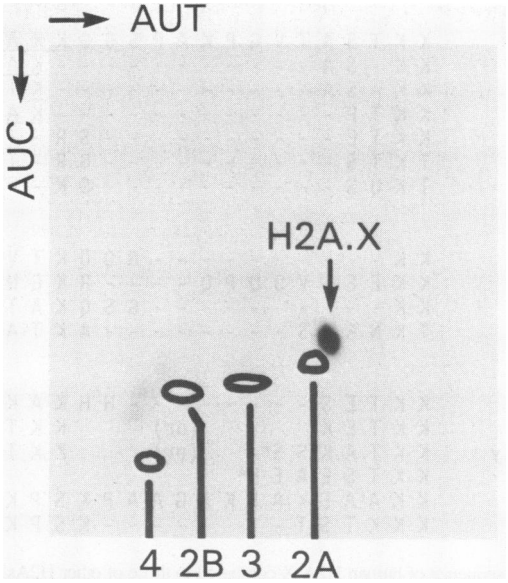

Figure 2. Coupled cell free transcription and translation of H2A.X cDNA. *In vitro* transcription and translation were performed as described in Methods and in Hatch and Bonner (22). The positions of added core histone standards on the two dimensional gel are shown. H2A.X standard was also present (but not shown for clarity) and comigrated with the labeled product of the cDNA (arrow).

at the carboxy-terminus. A similar carboxy-terminal tetrapeptide is also found in Drosophila H2AvD (32) with the acidic residue replaced by an alanine residue. In some of these species, the homology also includes several more neighboring amino acid residues. Amino acid residues 121 to 133, between the conserved core and carboxy-terminal motifs in the H2A.X sequence, do not seem to be homologous with the human H2A.1 or any other published H2A sequence, however, the DNA sequences in this region of these two human H2As do seem to be related (discussed below).

H2A.X, like other mammalian H2As, is partially ubiquitinated (1). In H2A.1, the site of ubiquitin attachment is the lysine at position 119 (40), and almost all sequenced H2As have adjacent lysines at positions 118 and 119. The H2A.X sequence has these adjacent lysines, but it also has another pair of adjacent lysines at positions 133 and 134. Prior peptide analysis of H2A.X (5) had showed that one lysine containing peptide was altered in migration between H2A.X and ubiquitinated H2A.X (uH2A.X), suggesting that H2A.X has one and only one ubiquitin attachment site. The migration of the altered peptide was consistent with it being composed of H2A.X residues 119 to 127 (41). It has also been shown that histones can be polyubiquitinated (42), but these are likely to be polymers of ubiquitin attached to one H2A site (43). We conclude from these studies that the attachment site for ubiquitin in H2A.X is likely to be the same as in H2A.1 and that the other pair of adjacent lysine residues at positions 133 and 134 is probably not involved in H2A.X ubiquitination.

H2A.X mRNA

Figure 1 also shows the complete nucleotide sequence of the human H2A.X cDNA. The total length of the cDNA is 1585 nucleotides, not counting 38 A residues in the polyA

```
H2A.X-type
Human H2A.X              K K T S A T V G P K A P S G G K K A T Q A S Q E Y*
Sac. cere H2A.1          K K - S A - - - - - - - - - - K A T K A S Q E L*
Sac. cere H2A.2          K K - S A - - - - - - - - - - K T A K A S Q E L*
Aspergillus H2A          K K T P - - - - - - - - - - - K A G K G S Q E L*
Tetrahymena H2A.1        K K T E - - - - - - - - - - S R - - G Q A S Q D I*
Schiz. pombe H2A.1       T K T S - - - - - - - - - - G R - T G K P S Q E L*
Schiz. pombe H2A.2       T K Q S - - - - - - - - - - G K - - G K P S Q E L*

H2A.F/Z-type
Human H2A.Z              K K - - - - - - - - - - G Q Q K T V*
Drosophila H2AvD         K K E E T V Q D P Q - - - - R K G N V I L S Q A Y*
Sea urchin H2A.F/Z       K K - - - - - - - - - - G S Q K A T*
Tetrahymena hv1          T K N R S S - - - - - - - A K T A E P R*

H2A.1/2-type
Human H2A.1 or .2        K K T E S - - - - - - - - H H K A K G K*
Drosophila H2A.1         K K T E K*        (or)       K K T E K*
Sea urchin H2A early     K K T A K S S*    (or)       K K T A K S S*
Tetrahymena H2A.2        K K T S E A E H*
Wheat germ H2A.2         K K A A E K A E K A G A A P K S P K K T T K S P K K A*
Wheat germ H2A.1         K K K T S T - - - - - - - - K S P K K K V A A K E*
```

Figure 3. Carboxy-terminal sequence of human H2A.X compared to those of other H2As. The amino acid sequence of human H2A.X beginning at lysine residue 118 is aligned with the same region of other H2As. Potential homologies are underlined. Dashes denote gaps to maximize alignments and do not denote missing amino acids. Asterisks denote the carboxy termini. H2A.X-type denotes H2A histone sequences with the highest homology to the human H2A.X sequence. H2A.F/Z-type and H2A.1/2-type denotes H2A histone sequences closely related to H2A.Z and H2A.1 respectively. These two types are classified on the basis of core amino acid sequence homologies. Sequences are from Saccharomyces cerevisiae (27), Aspergillus nidulans (28), Tetrahymena H2A.1 and .2 (31), Schizosaccharomyces pombe (29,30), Drosophila H2AvD (32), Tetrahymena hv1 (33), Strongylocentrotus purpuratus (sea urchin) H2A.F/Z (34), Human H2A.Z (22), Drosophila H2A.1 (35), Strongylocentrotus purpuratus H2A early (36), Human H2A.1 (see Methods,21,37), Triticum vulgare H2A.1 (38) and .2 (39)

tail of this particular isolate. There are 73 nucleotides in the 5′ UTR, 432 in the coding region, and 1080 in the 3′ UTR. The cDNA has an overall GC content of 67% with the expected chance frequency of CpG dinucleotides, suggesting that the gene is present in a CpG island in the chromatin (44,45).

The 5′ UTR contains 73 bases with a 59% GC content. Figure 4 shows by primer extension analysis that the major start site of the H2A.X mRNA seems to be at the C residue at position 2 of the sequence of H2A.X clone 12a; however, there are two minor bands (noted in Figure 4) just in the G-linker of the plasmid. This result suggests that the mRNA molecule yielding cDNA clone 12a may have had one or two G residues at the 5′ end, these being indistinguishable from the G-linker. This 5′ end yields lengths consistent with the measured lengths of both forms of the H2A.X mRNA (discussed below and Figure 6).

The 5′ UTR does not seem to be unusual (46). Its translation initiation sequence GCtAgCATG is similar to the consensus sequence GCC(AG)CCATG (differences with the consensus sequence are in lower case) for vertebrate genes. It does not contain any upstream initiation codons like 90% of vertebrate genes. Its length falls in the common size range of 20−100 bases. It has no recognizable homology with the 5′ UTR of the H2A.1 mRNA.

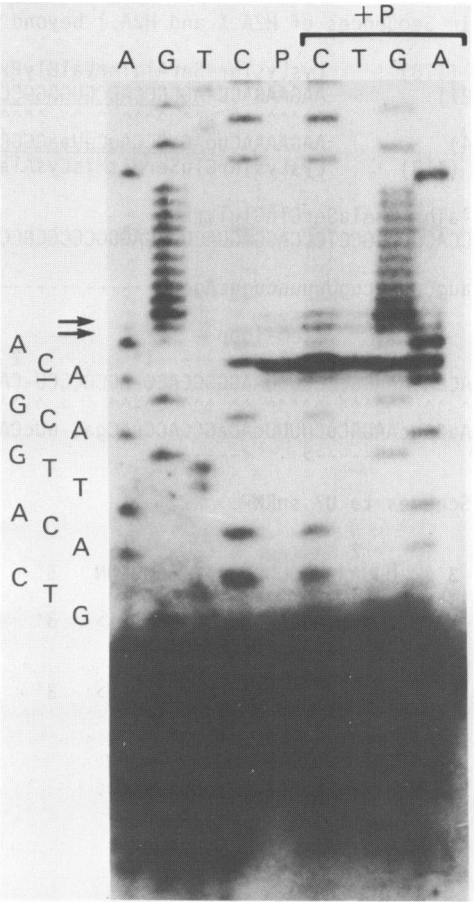

Figure 4. Primer extension of the H2A.X mRNA. The procedure is described in Methods. The primer extended on total RNA from proliferating Jurkat cell cultures was loaded onto lane P. The four lanes on the left show the four sequencing reactions using the same primer on the cDNA. The four lanes on the right are the same sequencing reactions mixed with the extended primer. Sequencing lanes are labeled so that reading down yields the transcript sequence. The ladder of 11 G residues is the linker to the Okayama-Berg vector, but would include any 5′ G residues in the cDNA.

The coding region is 70% GC, and 12% T. Most of the T residues are in the second coding register which has the least flexibility due to degeneracy; this register is 49% GC and 23% T, while the first coding register, which has some degeneracy, is 69% GC and 8% T. However, the 3rd coding register is 92% GC. This preference for GC-rich codons, although noted for several histone DNA sequences (47), is quite marked for H2A.X with 131 of 142 codons having the maximum possible GC content, compared to H2A.1 with 89 of 129 and H2A.Z with 61 of 127. Between H2A.X and H2A.1, the overall homology of the first 363 bases of the coding region is 81%. There is a stretch of 38 bases from

A. RNA and Protein Sequences of H2A.X and H2A.1 beyond Lysine$_{118}$.

```
(XP) H2A.X Protein (118)      LysLysThr-SerAlaThrValGlyProLysAlaProSer >
(XR) H2A.X cDNA (427)         AAGAAGACC-AGCGCCACCGUGGGGCCGAAGGCGCCCUCG >
                              ^^^^^ ^^ _^^ ^^^^ ^ ^  ^^^^ ^ ^^
(1R) H2A.1 DNA (404)          AAGAAAACugAGaGCCAuCaUaaGGCCaAgGGaaagugaa >
(1P) H2A.1 Protein (118)      LysLysThrGluSerHisHisLysAlaLysGlyLysEND

(XP) GlyGlyLysLysAlaThrGlnAlaSerGlnGluTyrEND
(XR) GGCGGCAAGAAGGCCACCCAGGCCTCCCAGGAGUACUAAGAGGGCCCGCGCCGCGGCCGGCCGCCCC >
     ^    ^ ^       ^ ^ ^   ^          ^  ^
(1R) GaguuaAcGcuucaugCaCuGcugUuuuucugucAgcAg--------------------------- >

                              stem-loop                    U7 binding
                          1 ------>  <------  1*        2           2
(XR) AGCUCCCCAUGCCACCACAAAGGCCCUUUUAAGGGCCACC-ACCGCCCU-CAUGGAAA-GAGCUG >
     ^      ^^  ^ ^ ^^  ^^^^^ ^^^^^ ^^ ^^^^^^-^^   _ ^-^^^ -^^^-^^^^^^
(1R) AcaaaauCAgcCuAaCAgcAAGGCuCUUUUcAGaGCCACCuACgaC-uUcCAUu-AAAuGAGCUG >
                         ----->    <-----
```

B. Base Pairing Schemes to U7 snRNP.

```
                          2          2
Human U7        3' < UUUAAGAUUUUCUCGACAUUGUGNN   5'

Human H2A.X     5'   CCUCAUGGAAAGAGCUGAGCCGC >   3'
                        1 1222323323 1 313

Human H2A.1     5'   UUCCAUUAAAUGAGCUGUUGUGC >   3'
                        12222 3233232 1113
```

C. Base Pairing Schemes of the Stem-Loop Dyads.

```
     U U                U U
    U   a              U   c
     U-A                U-A
     C-G                C-G
     c-g                u-a
     C-G                C-G
     G-C                G-C
   AAG-CACC           AAG-CACC
552      573
   H2A.X               H2A.1
```

Figure 5. Sequence comparison between H2A.X and H2A.1. A). RNA and protein sequences of H2A.X and H2A.1 beginning with lysine$_{118}$. A caret denotes a matching base, a dash denotes a gap to maximize alignment, mismatched bases in H2A.1 are shown in lower case. The underlined RNA sequences just past threonine$_{120}$ show their homology after a shift in the reading frame. The conserved elememts, 1 and 2 are as in Figure 1. The asterisk just after the stem-loop is the presumed site of cleavage for the short form of H2A.X mRNA. B). Base pairing homologies between the human U7 snRNP and conserved motif 2 in H2A.X and H2A.1. Potential bonds are noted by a 3 for a GC pair, a 2 for an AU pair, and a 1 for a GU pair. C). Base-pairing structures for the stem-loop elements of H2A.X and H2A.1 mRNAs. Since energy considerations indicate that loops with less than three bases will not form (67,68), the H2A.X stem-loop structure is shown with a four base loop.

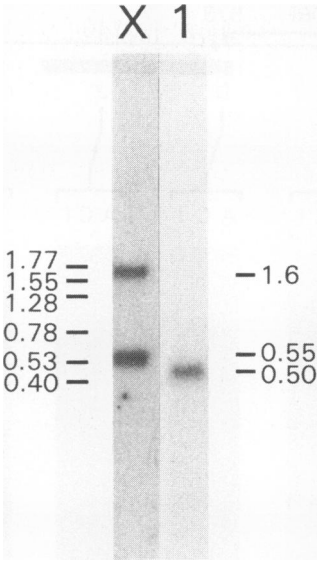

Figure 6. Length of H2A.X mRNAs. Total RNA from proliferating Jurkat cell cultures was electrophoresed in parallel with RNA standards (BRL) and electroblotted onto Zetaprobe. The membrane was first probed with an H2A.X specific probe (dyad probe, see Figure 7), then reprobed with the H2A.1 gene fragment. Standards on the membrane were visualized with methylene blue.

H2A.X positions 266 to 303 that is identical in sequence with that of the human H2A.1 gene used to select this clone.

As previously discussed, the protein sequences of H2A.X and H2A.1 past Thr_{120} are not homologous; however, the RNA sequences in this region are related (underlined sequences in Figure 5A). An insertion of a G residue into the H2A.X sequence directly after Thr_{120} increases the alignment with the H2A.1 sequence in this region to 68% positional identity, much more than that obtained without this frameshift. This finding suggests an evolutionary relationship between the two DNA sequences, even though the protein sequences seem unrelated.

H2A.X mRNA 3' UTR

The H2A.X mRNA 3' UTR is unusual in that in contains not only polyA addition motifs and polyA shown to be involved in replication-independent type histone mRNAs (48,49) but also stem-loop and U7 binding sequences shown to be involved in the processing and stability of replication-type histone mRNAs (8–12; Figure 1).

There is a polyadenylation motif, ATTAAA (10), 16 bases upstream from the polyA tail (38 A residues in clone 12a). Downstream from the polyadenylation motif is a T-rich element similar to those that have been found to play a role in the efficient 3' polyadenylation of other poly A mRNAs (10). However, just downstream from the translation termination signal are also highly conserved motifs found in replication-dependent histone genes (50,51). Figure 5 shows the conservation of these sequences between the H2A.X and H2A.1 mRNAs (panel A), the base pairing scheme of these two sequences with human U7 snRNP (52; panel B), and possible stem-loop structures in this region (panel C).

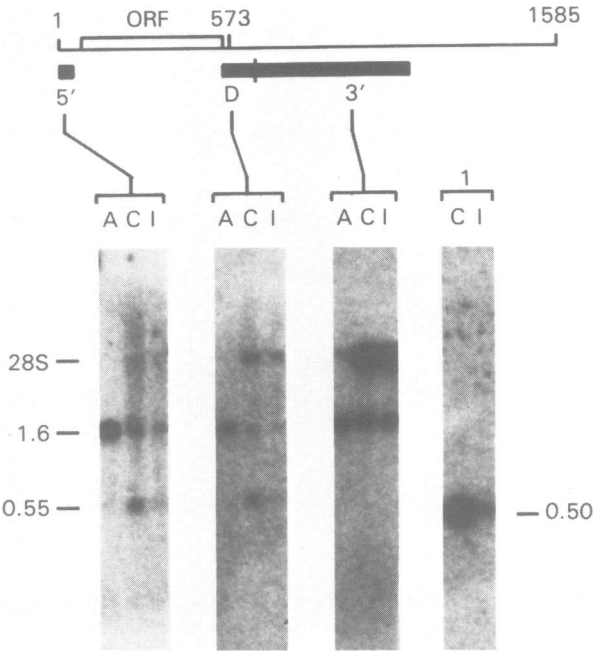

Figure 7. H2A.X mRNA alternative processing. RNA samples used were; (C) total RNA from proliferating Jurkat cell cultures, (I) total RNA from proliferating Jurkat cell cultures preincubated with 10 mM hydroxyurea, and (A) polyA selected RNA from proliferating Jurkat cell cultures. The RNA samples were electrophoresed and electroblotted onto Zetaprobe as described in Methods. The probes used were; (5′), from bases 1−60 of the 5′ UTR, (D or dyad), bases 521−630 of the 3′ UTR containing 54 bases 5′ and 57 bases 3′ of the presumed processing site at base 573, (3′), bases 631−1123 of the 3′ UTR, and (I) the H2A.1 gene fragment. The dyad probe and particularly the 3′ probe crosshybridize to 28S RNA. The 3′ probe contains GC runs which were found to have high homology with sequences in the human 28S rRNA (data not presented).

The region encompassing the final 14 amino acids of the H2A.X ORF does not have any recognizable homology with the H2A.1 mRNA (Figure 5A). Just beyond the end of the H2A.X mRNA ORF, there is a 26 base GC sequence which corresponds closely to the difference in distance between the conserved protein coding regions and the conserved dyad and U7 binding motifs in these two mRNAs. In the H2A.X mRNA there are 49 bases between the end of the ORF and the beginning of the stem-loop motif compared to 61 for the H2A.1 mRNA. Just 5′ to the dyad is the eight nucleotide sequence ACCACAAA reported to be highly conserved in replication type histone mRNAs of vertebrates (9). The H2A.X mRNA stem-loop sequence is 16 bases long, the same length as those of other replication type histone mRNAs. The H2A.X U7 binding sequence (Figure 5B) is complementary to the U7 sequence for 9 contiguous nucleotides (10 counting a GU pair at one end), while the H2A.1 U7 binding sequence has a mismatch between runs of 4 (5 counting a GU pair) and 7 nucleotides.

Two forms of H2A.X mRNA

H2A.X is considered a basal histone since it is synthesized in G1 as well as in S-phase (2,3). Quantitative studies (2,3) suggested that there is a 3−7 fold increase in H2A.X

synthesis during S-phase relative to that in G1, compared to a 60 fold increase for H2A.1. When total RNA from exponentially rowing Jurkat cell cultures was hybridized with a restriction fragment containing sequences upstream from the replication type processing site at nucleotides 573−574 (the dyad probe in Figure 7), two lengths of RNA were found (Figure 6). The longer mRNA was measured against standard RNAs to be 1600 bases long, a length consistent with that of the cDNA sequence. The shorter mRNA was measured at 550 bases long, a length consistent with the 573 bases long mRNA expected from the cDNA sequence. When the same blot was hybridized with the replication-linked H2A.1 probe, its mRNA was measured to be slightly shorter at 500 bases, consistent with its expected length of 528 bases (21). Thus, the H2A.X mRNA was measured to be about 50 bases longer than that of the H2A.1 mRNA, a difference which agrees with the 46 base difference predicted from their sequences.

Further studies were performed with polyA selected RNA and with total RNA from cell cultures that had been preincubated with inhibitors of DNA synthesis (Figure 7). With the 5′ and dyad probes, two probes containing sequences upstream from the presumed processing site at base 573, the shorter mRNA was found to be sensitive to the inhibition of DNA synthesis. In these cells under control conditions, there seems to be about twice as much short as long H2A.X mRNA. The short H2A.X mRNA was absent from polyA selected RNA; in addition, a restriction fragment containing only sequences downstream (bases 631−1123) from base 573 visualized only the longer mRNA. These results indicate that the shorter mRNA seems to behave as a typical replication-type histone mRNA, while the longer as a replication-independent type histone mRNA. It is striking that these two forms of H2A.X mRNA seem to coexist in the same cell culture and perhaps simultaneously in the same cells.

DISCUSSION

This study presents the first complete amino acid sequence of an H2A.X isoprotein species. The H2A.X sequence is almost identical to that of human H2A.1 up to amino acid residue 120. However the carboxy-terminal region beyond residue 120 is not homologous to any other sequenced vertebrate H2A (25), but it is similar to the carboxy-terminal regions of several H2As from lower eukaryotes (26−39; Figure 3).

Generally, histone H2A sequences have a highly conserved region from residues 31 to 95, containing the central core of the H2A protein, considered to be the site of interaction with H2B to form the H2A:H2B dimer (53). This core is flanked by amino- and carboxy-terminal sequences that seem to be much less conserved. Because the H2A carboxy-terminal sequences are generally quite variable (25,54), the homology between these carboxy-terminal sequences is all the more striking, particularly considering the genetic distance between these various organisms. In the highly conserved H2A core region, there are 2−3 sequence differences between mammals and two of the yeasts, A. nidulans and S. pombe, which have identical sequences in this region. Between mammals and S. cere., there are 7 differences; and 15 between mammals and Tetrahymena. However, between Tetrahymena and the three yeasts, there are also about 15 sequence differences, and between S. cere. and the other two yeasts, there are 5. Thus in evolutionary terms, except for A. nidulans and S. pombe, these other species have diverged more from each other than mammals have from S. pombe.

There are several possible explanations for these carboxy-terminal sequence homologies. First, it is possible that they arose by coincidence, but this is probably unlikely as there

are 7−8 sequences from diverse species with the same motif. A second possibility related to the first is that these homologies arose by convergent evolution, with this type of sequence being part of a small number of carboxy-terminal sequences consistent with H2A function. If this were the case, one would expect other conserved motifs to be apparent in the other H2A-type sequences, but those sequences are more variable and quite different from the X-type.

A third possibility is that this type of carboxy-terminal sequence has been maintained during evolution because it participates in an essential cellular function. The more divergent H2A.Z isoprotein species, with many amino acid differences in the H2A core region has been shown to be conserved throughout animal evolution as separate H2A species (55,56), presumably because it performs an essential function that can not be performed by any other H2A isoprotein. Perhaps the H2A.X carboxy-terminal sequence motif has also been conserved during evolution, but has escaped detection in multicellular animals because most of the H2A.X sequence is almost identical with that of the much more plentiful H2A.1/2 species. In lower eukaryotes an H2A.X type might be a larger fraction of the H2A complement. In the yeasts, both H2A isoproteins carry this motif. Tetrahymena contains an H2A species of each of the three types. As a further speculation, since it seems to be the carboxy-terminal sequence motif that distinguishes the H2A.X type, perhaps this motif could also be attached to an H2A.F/Z type; as may be the case with Drosophila H2AvD. However, if this carboxy-terminal sequence motif is essential, then it should be found on at least one H2A species in each organism throughout evolution; the validity of such a hypothesis must await further studies.

Yeast chromatin is composed largely of transcribable sequences; if the carboxy-terminal region of H2A.X does have some function related to active chromatin, then yeast H2As might be expected to be more like H2A.X than H2A.1 isoproteins. In this regard, it is also interesting to note that Huang et al. (57) have reported evidence that the transcriptionally active mouse immunoglobulin kappa chain gene seems to be packaged in nucleosomes enriched in H2A.X.

Replication-dependent histone genes contain conserved stem-loop and U7 binding sequence motifs (10,11,50,51; H2A.1 motifs 1 and 2 in Figure 5) which are involved in the processing and stability of the mRNA (8−12,58−61). Mature replication-dependent histone mRNAs have been processed so that they terminate just past this stem-loop motif (asterisk in Figure 5A). The mRNA of the other basal H2A isoprotein species, H2A.Z, lacks the stem-loop structure (22,32−34,62), as do the mRNAs of other sequenced basal histone species such as H3.3 (48) and a germline H2B (63). In contrast, the H2A.X mRNA contains sequences highly homologous to both motifs found in replication-dependent histone genes in addition to the polyA motifs. The two H2A.X mRNAs are consistent in size and hybridization characteristics with these two alternative processing pathways. Recently, Challoner et al. (64) have reported that the mRNA of an avian H2B gene can exist in two sizes due to alternative 3′ end formation. In somatic cells a 0.5 kb mRNA was found, but in spermatid cells, the mRNA was 0.8 kb long and contained a polyA tail begining 26−28 nucleotides 3′ of the predicted terminus of a replication type histone mRNA. Evidence for a similar situation was also found for some H2A and H3 genes. A similar situation has been reported in Xenopus oocytes (65,66) where a polyA tail is found attached immediately 3′ to the conventional site of 3′ cleavage. Thus, the H2A.X gene seems to be a member of the family of histone genes that transcribe mRNAs with more than one pathway for processing 3′ ends.

The H2A.X mRNA differs from these other cases in two ways. First, the long H2A.X mRNA continues for over 1000 bases before the polyA tail. Second, in contrast to these other reports in which the long mRNA seems to be present only in germline cells, both H2A.X mRNA lengths are found in somatic cells, and in addition both lengths are found in the same proliferating cell culture. Previous results from this laboratory suggested that the rate of synthesis of H2A.X protein may increase about 3−7 fold in cells entering S-phase, much less than the 60 fold seen for H2A.1, but more than that for H2A.Z (2). It is possible that the 1.6 kb mRNA is formed only in cells not in S-phase and the 0.55 kb mRNA only in S-phase cells. However, because synchronized S-phase cells with inhibited replication were still found to synthesize quite significant amounts of H2A.X (2), we suspect that the 1.6 kb as well as the 0.55 kb mRNAs may be both present in S-phase cells.

The basal library also contains several other H2A clones. One of these, H2A.1 clone 6a, has been sufficiently sequenced to show that it, like H2A.X, contains both replication type and basal type processing signals (unpublished observations). These results indicate that the H2A.X mRNA is not unique in somatic cells, and that alternative 3′ end formation of histone mRNAs may be involved in somatic cells in G1 and G0 when DNA replication is absent. One intriguing question raised by these studies then is why are the replication type signals maintained in these mRNAs, if the histones they encode, like H2A.X, is only 10−15% of the H2A complement in mammalian tissues. We are investigating the expression of this and similar genes during the cell cycle of proliferating cells and in quiescent cells.

*To whom correspondence should be addressed

REFERENCES

1. West,M.H.P., and Bonner,W.M. (1980) Biochemistry *19*, 3238−3245.
2. u,R.S., and Bonner,W.M. (1981) Cell *27*, 321−330.
3. u,R.S., Tsai,S., and Bonner,W.M. (1982) Cell *31*, 367−374.
4. weidler,A. (1976) Life Sci. Res. Rep. *4*, 187−196.
5. West,M.H.P., and Bonner,W.M. (1983) Comp. Biochem. Physiol. *76B*, 455−464.
6. onner,W.M., West,M.H.P., and Stedman,J.D. (1980) Eur. J. Biochem. *109*, 17−23.
7. Okayama,H., and Berg,P. (1983) Molec. Cell. Biol. *3*, 280−289.
8. Pandey,N.B., and Marzluff,W.F. (1987) Molec. Cell. Biol. *7*, 4557−4559.
9. Stauber,C., Luscher,B., Eckner,R., Lotscher,E., and Schumperli,D. (1986) EMBO J. *5*, 3297−3303.
10. Birnstiel,M.L., Busslinger,M., and Strub,K. (1985) Cell *41*, 349−359.
11. Schumperli,D. (1988) TIG *4*,187−191.
12. Marzluff,W.F., and Pandey,N.B. (1988) TIBS *13*, 49−52.
13. Sittman,D.B., Graves,R.A., and Marzluff,W.F. (1983) Proc. Natl. Acad. Sci. USA *80*, 1849−1853.
14. De Lisle,A.J., Graves,R.A., Marzluff,W.F., and Johnson,L.F. (1983) Mol. Cell. Biol. *3*, 1920−1929.
15. Heintz,N., Sive,H.L., and Roeder,R.G. (1983) Mol. Cell. Biol. *3*, 539−550.
16. Sanger,F., Nicklen,S., and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA *74*, 5463−5467.
17. Barnes,W.M., Bevan,M., and Son,P.H. (1983) Methods in Enzymology *101*, 98−122.
18. Mizusawa,S., Nishimura,S., and Seela,F. (1986) Nucleic Acid Res. *14*, 1319−1324.
19. Sequenase Protocol Booklet, (1988) United States Biochemical Corp.
20. Hatch,C.L., and Bonner,W.M. Unpublished data.
21. Zhong,R., Roeder,R.G., and Heintz,N. (1983) Nucleic Acid Res. *11*, 7409−7425.
22. Hatch,C.L., and Bonner,W.M. (1988) Nucleic Acid Res. *16*, 1113−1124.
23. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.A., and Struhl,K.C in Current Protocol in Molecular Biology (1989) Green Publishing Associates and Wiley-Interscience.
24. Maniatis,Y., Fritsch,E.F., and Sambrook,J. (1982) Cold Spring Harbor Laboratory,Cold Spring Harbor,N.Y.
25. Wells,D.E. and McBride, C. (1989) Nucleic Acid Res. *17*, r311−r346.
26. Conner,W., Mezquita,J., Winkfein,R.J., States,J.C., and Dixon,G.H. (1984) J. Mol. Evol. *20*, 227−235.
27. Choe,J., Kolodrubetz,D., and Grunstein,M. (1982) Proc. Natl. Acad. Sci. USA *79*, 1484−1487.

28. May,G.S., and Morris,N.R. (1987) Gene *58*, 59–66.
29. Matsumoto,S., and Yanagida,M. (1985) EMBO J. *4*, 3531–3538.
30. Choe,J., Schuster,T., and Grunstein,M. (1985) Molec. Cell. Biol. *5*, 3261–3269.
31. Fusauchi,Y., and Iwai,K. (1983) J. Biochem. *93*, 1487–1497.
32. Van Daal,A., White,E.M., Gorovsky,M.A., and Elgin,C.R. (1988) Nucleic Acids Res. *16*, 7487–7497.
33. White,E.M., Shapiro,D.L., Allis,C.D., and Gorovsky,M.A. (1988) Nucleic Acids Res. *16*, 179–198.
34. Ernst,S.G., Miller,H., Brenner,C.A., Nocente-McGrath,C., Francis,S., and McIsaac,R. (1987) Nucleic Acids Res. *15*, 4629–4644.
35. Goldberg,M.L. (1979) PhD. Thesis 'Sequence Analysis of Drosophila Histone Genes', Stanford University.
36. Sures,I., Lowry,J., and Kedes,L.H. (1978) Cell *15*, 1033–1044.
37. Hayashi,T., Ohe,Y., Hayashi,H., and Iwai,K. (1980) J. Biochem. *88*, 27–34.
38. Rodrigues,J.A., Brandt,W.F., and von Holt,C. (1985) Eur. J. Biochem. *150*, 499–506.
39. Rodrigues,J.A., Brandt,W.F., and von Holt,C. (1988) Eur. J. Biochem. *173*, 555–560.
40. Goldknopf,I.L., and Busch,H. (1977) Proc. Natl. Acad. Sci. USA *74*, 864–868.
41. West,M.H.P., Wu,R.S., and Bonner,W.M. (1984) Electrophoresis *5*, 133–138.
42. Davie,J.R., Delcuve,G.P., Nickel,B.E., Moyer,R., and Bailey,G. (1987) Cancer Res. *47*, 5407–5410.
43. Nickel,B.E., and Davie,J.D. (1989) Biochemistry *28*, 964–968.
44. Bird,A.P. (1986) Nature (London) *321*, 209–213.
45. Gardiner-Garden,M., and Frommer,M. (1987) J. Mol. Biol. *196*, 261–282.
46. Kozak,M. (1987) Nucleic Acids Res. *15*, 8125–8148.
47. Sittman,D.B., Graves,R.A.,and Marzluff,W.F. (1983) Nucleic Acid Res. *11*, 6679–6697.
48. Wells,D., and Kedes,L. (1985) Proc. Natl. Acad. Sci. USA *82*, 2834–2838.
49. Bird,R.C., Jacobs,F.A., Stein,G., Stein,J., Sells,B.H., (1985) Proc. Natl. Acad. Sci. USA *82*, 6760–6764.
50. Busslinger,M., Portmann,R., and Birnstiel,M.L. (1979) Nucleic Acids Res. *6*, 2997–3008.
51. Hentschel,C.C., and Birnstiel,M.L. (1981) Cell *25*, 301–313.
52. Mowry,K.L. and Steitz,J.A. (1987) Science *238*,1682–1687.
53. Moss,T., Cry,P.D., Aberchrombie,B.D., Crane-Robinson,C., and Bradbury, E.M. (1976) Eur. J. Biochem. *71*, 337–350.
54. Wu,R.S., Panusz,H.T., Hatch,C.L., and Bonner,W.M. (1986) Crit. Rev. in Biochem. *20*, 201–263.
55. Wu,R.S., Nishiokia,D., and Bonner,W.M. (1982) J. Cell. Biol. *93*, 426–431.
56. Mueller,R.D., Yasuda,H., Hatch,C.L., Bonner,W.M., and Bradbury,E.M. (1985) J. Biol. Chem. *260*, 5147–5153.
57. Huang,S.Y., Barnard,M.B., Xu,M., Matsui,S.I., Rose,S.M., and Garrard, W.T. (1986) Proc. Natl. Acad. Sci. USA *83*, 3738–3742.
58. Seiler-Tuyns,A., and Paterson,B.M. (1986) Nucleic Acids Res. *14*, 8845–8862.
59. Ross,J., Peltz,S.W., Kobs,G., and Brewer,G. (1986) Mol.Cell.Biol. *6*, 4362–4371.
60. Ross,J. and Kobs,G. (1986) J.Mol.Biol. *188*, 579–593.
61. Gick,O., Kramer,A., Vasserot,A., and Birnstiel,M.L. (1987) Proc. Natl. Acad. Sci. USA *84*, 8937–8940.
62. Harvey,R.P., Whiting,J.A., Coles,L.S., Krieg,P.A., and Wells,J.R.E. (1983) Proc. Natl. Acad. Sci. USA *80*, 2819–2823.
63. Moss,S.B., Challoner,P.B., and Groudine,M. (1989) Dev. Biol. *133*, 83–92.
64. Challoner,P.B., Moss,S.B., and Groudine,M. (1989) Mol.Cell.Biol. *9*, 902–913.
65. Ballantine,J.E.M. and Woodland,H.R. (1985) Febs Letters, *180*, 224–228.
66. Chambers,A. and Old,R. (1988) Dev.Biol. *125*, 237–245.
67. Salser, W. (1977) Cold Spring Harbor Symp. Quant. Biol. *42*, 985–1002.
68. Gouy,M. (1987) in Bishop,M.J. and Rawlings,C.J. (ed.),Nucleic Acid and Protein Sequence Analysis-A Practical Approach. IRL Press, Oxford, pp 259–284.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.