Structure and in vitro transcription of a human H4 histone gene

F. Sierra, G. Stein and J. Stein

University of Florida, College of Medicine, Gainesville, FL 32610, USA

Received 2 June 1983; Revised and Accepted 20 September 1983

ABSTRACT

A human H4 histone gene was isolated and the nucleotide sequences of the mRNA coding as well as the 5' and 3' flanking regions were determined. No intervening sequences were found in this gene. A series of sequences which have been assigned putative regulatory roles in histone genes and/or in other genes were identified both upstream and downstream from the H4 histone protein coding region. Deletion mutants were constructed by BAL-31 nuclease digestion of sequences in the 5' flanking region of this H4 histone gene and were assayed in an in vitro transcription system. No regions upstream from the TATA box were required for site specific initiation in vitro. Data are presented which suggest that sequences located downstream from the 3' end of the coding region may influence the in vitro transcription of this human H4 histone gene.

INTRODUCTION

The histone genes of several vertebrates, including humans, are represented as a family of moderately repeated genes, arranged in clusters but there is no evidence of a simple tandem repeat organization (1-16, reviewed in 34). In contrast, the early histone genes from several sea urchin species, as well as those of Drosophila, appear to be clustered and tandemly repeated (17).

In several human cell lines, histone genes present in different clusters appear to be coordinately expressed during the cell cycle (18,19). For example, S phase HeLa cells contain at least seven different polysomal H4 mRNAs (20). These different H4 mRNA species appear to code for the same H4 protein, as detected by peptide mapping of their in vitro translation products (21). S1 nuclease protection studies have indicated that one of the major subspecies of histone H4 mRNA found in HeLa cells is encoded by a gene similar or identical to the H4 histone gene cloned in the λCh 4A genomic recombinant λHHG 41 (described in 15).

Both in vivo and in vitro transcription studies performed on a variety of genes (22-25) have implicated the TATA box as playing a role in directing the
precise site of initiation of transcription by RNA polymerase II. Other sequences upstream from the TATA box have been tentatively assigned promoter functions, based on their conservation in several related and unrelated genes. One of these sequences is the 5'-GGPyCAATCT-3', or "CAAT" box, described by Benoist et al. (26) and Efstratiadis et al. (22), which is found between 70 and 80 nucleotides upstream from the start site of many genes. Deletion of the "CAAT" box does not decrease the in vitro transcription capacity of any gene tested so far, including sea urchin histone H2A (23,25,27) conalbumin (27) and human and rabbit α-globin (24) genes. Sequences upstream from the "CAAT" box have been found to have an effect on the in vivo transcription of several genes (23,28-30); however, these sequences seem to have no effect on the in vitro transcription of most genes (24). An exception to this has been found in the case of a sea urchin H2A gene, where deletion of a region comprising nucleotides -111 to -139 (starting from the cap site) seems to produce a down mutation when assayed in vitro (25).

In this paper we present nucleotide sequence data indicating that the human H4 histone gene present in λHHG 41 encodes a typical H4 histone protein. We have constructed a series of 5' deletion mutants of this H4 gene and used them to examine the effect of several putative regulatory regions on in vitro transcription. The results indicate that sequences upstream from the TATA box are not required for in vitro transcription of this gene. Preliminary observations suggest that DNA sequences located downstream from the 3' end of this gene may be involved in the production and/or stability of in vitro run-off transcripts.

MATERIALS AND METHODS
Plasmid Growth and Purification
Plasmids pF0108 and pF0108A were prepared by insertion of restriction fragments derived from λHHG 41 (15) into pBR322 and were then transformed into E. coli strain HB101. Bacteria containing the recombinant plasmids were grown in L-broth in the presence of 50 μg/ml ampicillin. Plasmid amplification was induced by addition of chloramphenicol to a concentration of 200 μg/ml. DNA was isolated by the cleared lysate procedure (31), followed by chromatography through a 30 x 1.5 cm BioGel A-15 m column, developed with 10 mM Tris-HCl (pH 8.0) - 1 mM EDTA. Fractions containing plasmid DNA (Vo) were pooled and ethanol precipitated.

All experiments involving viable microorganisms containing recombinant DNA molecules were performed under conditions specified by the NIH Guidelines.
for Research Involving Recombinant DNA.

**Construction of 5' Deletion Mutants**

Ten µg of DNA from the plasmid pF0108A (Fig. 1) were digested to completion with EcoRI, phenol extracted and ethanol precipitated. Nucleotides were removed from the ends of the DNA by digestion with nuclease BAL-31 (BRL) in a 400 µl reaction, containing 20 mM Tris-HCl (pH 8.1)-200 mM NaCl-12 mM CaCl2-1 mM EDTA (32). The reaction was started by addition of 2.0 units of BAL-31 and incubated at 30°C for a total of 15 minutes. Aliquots were withdrawn every 30 seconds and pooled aliquots were frozen in the presence of 20 mM EDTA-20 mM EGTA. Deproteinized samples were then ligated to synthetic EcoRI linkers in the presence of T4 DNA ligase, followed by extensive digestion with EcoRI and HindIII. The samples were separated from the excess linkers by chromatography on a 9.5 x 0.9 cm BioGel A-15 column. The resulting DNA fragments were ligated to EcoRI/HindIII-digested pBR322 DNA that had previously been treated with calf intestine alkaline phosphatase. Transformation of E. coli strain HB101 was done by standard procedures.

**DNA Sequence Analysis**

The H4 histone gene present in pF0108A (Fig. 1) resides primarily within the 317 bp SacI I fragment, with its 5' end located within the adjacent 408 bp EcoRI/SacII fragment. DNA was digested with both restriction endonucleases, dephosphorylated with bacterial alkaline phosphatase and labeled at its 5' end by T4 polynucleotide kinase in the presence of [α-32p]-ATP. After strand separation (33), the DNA was sequenced by the chemical modification method of Maxam and Gilbert (33).

**In Vitro Transcription**

In vitro transcription was performed using the HeLa whole cell extract described by Manley et al. (34). Reactions were performed using 30 µl of lysate, 2.5 µg of DNA (50 µg/ml final), 1 µl of 7 mM EDTA (0.2 mM final), 1 µl of 50 mM phosphocreatine (1 mM final), 5 µl of 10 mM NTP (1 mM final of ATP, GTP, and CTP, 0.05 mM UTP) and 20 µCi of [α-32p] UTP in a total volume of 50 µl. Transcription was allowed to proceed for 50 minutes at 30°C. At this time, cold UTP was added to a concentration of 1 mM, and incubation was continued for 15 minutes at 30°C, to chase partially synthesized, labeled RNA molecules into full-size transcripts.

**Isolation of In Vitro Transcription Products**

In vitro transcription reactions were terminated by addition of 55 µl of 10% SDS and 195 µl of fresh 2 mM Tris-HCl (pH 7.4)-1 mM EDTA-2 µg/ml polyvinyl sulfate-1 µg/µl proteinase K. The contents of the tube were vortexed gently.
and digestion by the protease was allowed to proceed for 15 minutes at room temperature. The solution was then adjusted to 0.25 M NaCl by addition of 15 μl of 5 M NaCl, and nucleic acids were isolated by one extraction with phenol:chloroform:isoamyl alcohol (25:24:1), followed by one extraction with chloroform:isoamyl alcohol (24:1). After precipitation in dry ice for 15 minutes with 3 volumes of ethanol in the presence of 0.25 M LiCl, the nucleic acids were pelleted (15 minutes in a microcentrifuge) and resuspended in 150 μl of 0.2% SDS, followed by addition of 150 μl of 2 M ammonium acetate. Nucleic acids were precipitated again with 3 volumes of ethanol, and after centrifugation were washed once with 70% ethanol. Incorporated radioactivity was determined by direct Cerenkov counting (35).

Analysis of In Vitro Transcripts on Formaldehyde Gels

Samples were resuspended in 16 μl of 50% (v/v) formamide-6% (v/v) formaldehyde-50 mM Na2SO4-1 mM EDTA, heated at 70°C for 5 minutes and quick chilled. Eight μl of dye [50% (v/v) glycerol-50% (v/v) formamide-0.05% bromo phenol blue (BPB)-0.05% xylene cyanole FF (XC)] were then added, and the sample was applied to a preelectrophoresed (30 minutes at 45 mA) 1.5% agarose-6% (v/v) formaldehyde gel. Electrophoresis was performed at 50 mA (about 65 V) for 3.5-4 hours in sample buffer minus formamide and containing only 3% formaldehyde (36). The gel was then dried and exposed to Kodak XAR-5 X-ray film, in the presence of an intensifying screen.

Primer Extension Analysis

DNA Primer. The DNA used as primer was a 64 bp Alul/HhaI fragment from pF0108A, containing sequences encoding amino acids 17 (Arg) to 38 (Ala) of the H4 protein. pF0108A DNA was digested with Alul, followed by phosphatase treatment and labeling in the presence of [α-32p] ATP and T4 polynucleotide kinase. The DNA was then digested with an excess of HhaI (5 units/μg of DNA) and the 64 bp fragment was isolated from a 10% polyacrylamide gel by electroelution (37).

RNA Samples. RNAs transcribed in vitro using 10 μg of restriction endonuclease-digested pF0108A DNA as a template were isolated as previously described. After the last wash with 70% ethanol, samples were resuspended in 500 μl of 10 mM Tris-HCl (pH 7.5)-2 mM CaCl2-10 mM MgCl2-2 μg/μl polyvinyl sulfate, and heated at 100°C for 5 minutes. After addition of 250 μg of yeast tRNA as carrier, the DNA template was digested with 25 μg of RNase-free DNase I for 10 minutes at 37°C. The sample was made 10 mM in EDTA, phenol extracted, chloroform:isoamyl alcohol (24:1) extracted, and ethanol precipitated in the presence of 0.25 M LiCl. RNase-free DNase I was prepared by
pre-incubating a 1 ug/ml solution of DNase I in 20 mM Tris-HCl (pH 8.0)-10 mM CaCl2 at 37°C for 20 minutes, followed by addition of proteinase K in the same buffer to a final concentration of 1 mg/ml. Incubation was for 2 hours at 37°C, after which the DNase was used directly (38).

Hybridization and Primer Extension. DNA fragments, in vitro transcription products and control HeLa cell polysomal 7-11S RNA were resuspended in 0.1 M NaCl. DNA and RNA were mixed in a total volume of 20 μl, denatured by heating at 100°C for 5 minutes, and transferred quickly to a 60°C water bath (39). Incubation was continued for 60 minutes, after which the water bath was turned off, thus allowing the samples to cool slowly to 40°C.

Primer extension was done in the same buffer as previously described (40) except that [α-32P] dCTP was omitted, and cold dCTP was added to a final concentration of 1 mM. The final volume was 50 μl. The reactions were started by addition of 10 units of AMV reverse transcriptase and allowed to proceed for 45 minutes at 37°C.

Analysis of Primer Extension Products. The RNA used as a template for reverse transcription was hydrolyzed by incubation in 0.1 M EDTA-0.2% SDS-0.3 M NaOH for 60 minutes at 50°C (41). The solution was then adjusted to 0.5 M Na acetate, phenol extracted, chloroform:isoamyl alcohol (24:1) extracted, and ethanol precipitated. Samples were analyzed in 10% polyacrylamide-8.3 M urea gels run for 6 hours at 17 W (27 mA). Urea was removed from the gel by two 15-minute washes in 50% ethanol and the gel was dried and exposed to X-ray film for autoradiography.

RESULTS

Sequence Analysis of a Human H4 Histone Gene

The 3.1 Kb EcoRI fragment from the human genomic clone XHdG 41 (15 and Fig. 1) has been subcloned into the EcoRI site of pBR322. This clone, pF0108 (Fig. 1), was mapped using several restriction endonucleases, and the position of the H4 gene was determined by hybridization to an H4 probe derived from the clone pCO2 (42), which contains a whole repeat of the early histone genes from the sea urchin Strongylocentrotus purpuratus. A smaller subclone, pF0108A (Fig. 1), was constructed by removal of the 1.3 Kb EcoRI/HindIII fragment containing the AluI sequence.

DNA sequence data from a 725 bp stretch of DNA, containing the entire H4 gene plus its flanking regions, are presented in Fig. 2. This sequence indicates that pF0108A DNA has the capacity to code for an H4 amino acid sequence identical to that of other H4 proteins, such as calf thymus (Fig. 2).
Figure 1: Restriction maps of XHHG 41 and subclones pFO108 and pFO108A. XHHG 41 is a λCh4a recombinant clone containing H3 and H4 human histone DNA sequences; its isolation and characterization have been described elsewhere (15). pFO108 contains the 3.1 Kb EcoRI fragment of XHHG 41 (containing the H4 histone gene) inserted into the EcoRI site of pBR322. pFO108 also contains a member of the Alu family of repeated DNA sequences, indicated by the white arrow. The 1.8 Kb EcoRI/HindIII fragment containing the H4 gene but not the Alu sequence has been subcloned into EcoRI/HindIII-digested pBR322 and is referred to as pFO108A. E = EcoRI, H = HindIII, B = BamHI, Sm = Smal, P = PstI, X = XbaI, Hc = HinclII, S = SacII.

Furthermore, the results suggest that in humans, as in most other species studied, histone-coding sequences (or at least this particular H4 gene) lack intervening sequences (43). The non-protein-coding region of the H4 gene also lacks intervening sequences, based on the location of putative regulatory regions and mRNA termination signals and on previous results from S1 nuclease protection experiments (20).

Sequences flanking the gene at its 5' end include several putative regulatory regions, such as a TATA box and a GGTCC motif which is similar to the GATCC motif found in an analogous position in several sea urchin histone genes (43). Further upstream are two CAAT boxes (indicated by closed boxes in Fig. 2), one of which is almost identical to those found in other genes transcribed by RNA polymerase II (22,26), including several histone genes (43). No such homology block has yet been found in the 5' flanking region of other H4 or H1 histone genes (43).

Further upstream from the H4 histone gene are several other non-random sequences. Notably, between nucleotides -152 and -174 from the ATG initiation
codon there is a stretch of 21 nucleotides which contains only A and G residues, most usually in the form of the trinucleotide GGA (Fig. 2). Similar, although not identical, sequences have been found in the spacer regions of other histone genes (17,43). Finally, several short direct repeats (indicated by horizontal arrows in Fig. 2) are present in the 5' flanking region of the H4 gene.

The 3' flanking region of the H4 gene present in pF0108A has been sequenced up to 107 nucleotides past the presumptive 3' end of the mRNA (Fig. 2). The sequence shows all the characteristic 3' features observed in other histone genes (43). For example, the mRNA is most likely terminated in vivo at the ACCA motif found a few nucleotides downstream from the hyphenated dyad symmetry which is characteristic of histone mRNA 3' ends (43). Twelve nucleotides downstream from the ACCA motif there is another histone gene-related motif, characterized mainly by its high A + G content. No specific function has yet been ascribed to these sequences.

**In Vitro Transcription of the H4 Histone Gene**

For characterization of these putative regulatory sequences, we optimized an in vitro transcription system based on the whole HeLa cell extract described by Manley et al. (43), for transcription of this specific human H4 histone gene. Figure 3 shows that, under appropriate conditions, EcoRI-digested pF0108 DNA can be used as a template in the in vitro transcription system, and a 2.9 Kb run-off transcript can be visualized on a 6% formaldehyde/1.5% agarose gel. Primer extension experiments have indicated that the transcripts are initiated primarily at the bona fide 5' end of the H4 histone gene (data not shown). Consistent with the observation of Detke et al. (44) that human histone genes are transcribed by RNA polymerase II in isolated nuclei, the production of the 2.9 Kb in vitro transcript is dependent on RNA polymerase II activity; no transcript of this size is observed when the in vitro transcription reaction is performed in the presence of 2 μg/ml of α-amanitin (Fig. 3). The α-amanitin-insensitive band located slightly above the 2.9 Kb transcript has also been observed when EcoRI-digested pBR322 DNA is used as a template.

When we attempted to truncate pF0108 DNA at positions within the H4 coding region, its ability to be used as a template in vitro was impaired, as measured by the run-off assay. These results will be described in a later section.

**The 5' Flanking Region**

The functional relevance of putative regulatory sequences in the 5'
Figure 2A: Nucleotide sequence data for a complete H4 gene and its flanking regions.

Sequences were obtained by the Maxam and Gilbert (33) method, using the strategy depicted in Fig. 2B. Small letters indicate residues that were not unequivocally assigned. An "N" indicates an undetermined nucleotide. Numbers above the sequence indicate the nucleotide position with respect to the A of the AUG initiation codon (designated 0) and decreasing towards the 5' terminus.

Boxes at the 5' end of the gene indicate the location of two tandem "CAAT" boxes, while boxes just preceding the 3' end of the gene indicate the TATA box and the purine-rich conserved sequences present in histone genes.

Horizontal arrows indicate short repeated sequences. Wavy underlines indicate further features present in the 5' flanking region of the H4 gene. These include the 21-nucleotide purine box between nucleotides -153 and -173, and the histone-related sequence GGTCC, located slightly upstream of the TATA box. The TATA box is located about 55 nucleotides preceding the ATG initiation codon.

Figure 2B: Sequencing strategy.

Restriction map of the H4 gene and the flanking regions present in λHHLG 41. Horizontal arrows below the map indicate the direction and extent of sequencing information obtained. The number of arrows below each region of the DNA indicates the number of times each fragment was sequenced.
flanking region of the H4 histone gene present in pF0108A was tested by constructing a series of deletion mutants which spanned the 5' flanking region. The clones were constructed by BAL-31 nuclease digestion of EcoRI-linearized pF0108A DNA (see Materials and Methods for details). The deletions present in these clones were sized by digesting DNA with the restriction endonucleases EcoRI and SacII, 3'-end labeling with the large fragment of E. coli DNA polymerase I, and sizing on 3% agarose (Fig. 4A) or 6% polyacrylamide gels (not shown). This digestion should produce a 317 bp fragment from every clone containing an H4 insert, plus an EcoRI/SacII fragment of varying size. The parental plasmid, pF0108A, produces an EcoRI/SacII fragment of 408 nucleotides in length. The third band observed in most clones present in Fig. 4A represents incompletely digested DNA. Figure 4B indicates the approximate location of the deletion endpoints, based on the size of the EcoRI/SacII fragment from each of the deletion mutants (Fig. 4A).

Several of these deletion mutants were tested in the in vitro transcription system after digestion with either HindIII or HindIII plus EcoRI. Figure 5 shows the in vitro transcripts obtained from EcoRI-HindIII digests of a representative sample of these clones. Numbers at the top of each lane indicate the approximate distance (in bp) between the endpoint of the deletion

---

**Figure 3:** In vitro transcription of EcoRI-digested pF0108 DNA. In vitro transcripts were analyzed in a 1.5% agarose, 6% formaldehyde gel. Transcripts were then visualized by autoradiography. Lane 1: In vitro transcripts synthesized using EcoRI-digested pF0108 as a template in the absence of α-amanitin; Lane 2: in the presence of 2 μg/ml of α-amanitin. The arrow on the left indicates the expected size of the H4 run-off transcript. Note that there is an α-amanitin insensitive transcript of higher molecular weight which is not an H4 histone gene transcript (arrow on right).
Figure 4A: Sizing of deletion mutants of pF0108A.

DNA from rapid preps was digested with EcoRI and SacII restriction endonucleases and radiolabeled with the Klenow fragment of E. coli DNA polymerase in the presence of [α-32P]dCTP. The labeled DNA fragments were resolved on a 3% agarose gel and were visualized by autoradiography. Numbers at the top of each lane indicate the approximate locations of the deletion endpoints with respect to the ATG initiation codon, as determined from this gel. Numbers at the left indicate the sizes (in base pairs) of pBR322/HinfI markers. The arrow indicates the position of the 317 bp SacII fragment of pF0108A.

Figure 4B: Endpoints of 5' deletion mutants of pF0108A.

Sequences upstream from the 5' end of the H4 coding region are shown. Vertical arrows indicate the approximate endpoints of the different deletion mutants, determined as described in the legend of Figure 4A. Nucleotide residues are numbered from the ATG initiation codon, and decrease in the upstream direction. The deletion clones are designated according to the nucleotide at which the endpoint of the deletion lies.
Figure 5: In vitro transcription of deletion mutants of pF0108A. Autoradiograms of 1.5% agarose/6% formaldehyde gels showing the transcripts obtained from selected deletion mutants. The template DNA was digested with EcoRI and HindIII restriction endonucleases. The arrows indicate the position of the expected run-off H4 transcripts. Numbers above the lines refer to the approximate positions of the endpoint with respect to the ATG initiation codon. The figure is a composite from the autoradiograms of two gels run back-to-back but represents samples from one experiment.

and the ATG initiation codon (see Fig. 2). While it is clear that the assay cannot be used in a quantitative manner, it is also obvious that all the deletions examined gave rise to an in vitro run-off transcript of the expected size (1.6 Kb), including clone pF0108A 5' A80, which is devoid of the direct repeats as well as the "CAAT" boxes previously described. These same deletion mutants exhibited a reduced level of in vitro transcription when the DNA template was digested with only HindIII, and not with EcoRI (not shown). Apparently EcoRI cleavage (5' to the H4 gene) can provide an artificial site of entry for the RNA polymerase, which then begins transcription at its normal site.

As was shown previously for the 2.9 Kb transcript from pF0108, the production of the 1.6 Kb in vitro transcript was sensitive to low concentrations of α-amanitin (4 μg/ml, data not shown). The accuracy of initiation of transcription from the deletion mutants was assayed by primer extension analysis. These results, shown in Fig. 6, indicate that when clone pF0108A 5' A80 is used as a template in the in vitro transcription system, the
transcripts are initiated at the same position as one of the in vivo synthesized H4 mRNAs from HeLa cells. We can therefore conclude that sequences more than 80 nucleotides upstream from the ATG initiation codon are not required for in vitro transcription of the H4 gene present in pF0108A. Similar results have been found for several other genes assayed in in vitro transcription systems (23,24,27), with only the TATA box being required for site-specific initiation of transcription.

The 3' Flanking Region

We have obtained results which suggest that sequences present downstream from the 3' end of the H4 gene may be involved in the production of run-off transcripts in vitro. When pF0108 DNA was truncated with the restriction enzymes EcoRI, SmaI, HindIII or PstI (see Fig. 1), a run-off transcript of the size expected in each case was observed (Fig. 7); however, if the template was truncated at positions close to the 3' end of or within the H4 histone gene (with restriction endonucleases XbaI, HincII or SacII), no α-amanitin-sensitive run-off transcript of the expected size was detected (Fig. 7). When the DNA was restricted with EcoRI + HincII, a transcript of the appropriate
Figure 7: In vitro transcription of pF0108 DNA digested with different restriction enzymes.

Autoradiograms of 1.5% agarose/6% formaldehyde gels showing in vitro transcripts obtained when using as a template pF0108 DNA digested with EcoRI plus a second restriction enzyme (indicated at the top of each lane). The black dot at the left of each lane indicates the position of the expected run-off H4 transcript. The open circle at the left of each lane indicates the position of a transcript derived from the Alu DNA sequence present in pF0108 (Fig. 1); this transcript is not observed when pF0108A is used as a template. Other transcripts are from the pBR322 vector DNA.

size was observed; however, this transcript was not α-amanitin sensitive and was also obtained when EcoRI/HincII-digested pBR322 DNA was transcribed in vitro (data not shown). Therefore, we concluded that it is not a transcript of the H4 gene present in pF0108. Analysis of these same in vitro transcription products in 5% polyacrylamide gels containing 8.3 M urea failed to identify smaller, specifically terminated H4 gene transcripts (data not shown). These results suggested that regions located downstream from the 3' end of the H4 histone gene might have an effect in enhancing in vitro initiation and/or elongation of transcription of the H4 histone gene present in pF0108.

To further explore these possibilities, a new series of primer extension experiments was executed to determine if initiation of transcription had
Figure 8: Primer extension analysis of 
in vitro transcripts. 
Autoradiograms of a 10% 
 polyacrylamide/8.3 M urea gel showing the 
 DNA obtained after extension of the 64 bp 
 AluI/HhaI primer by AMV reverse 
 transcriptase, using different RNA samples 
 as templates. Lane 1: HeLa polysomal RNA. 
 Lanes 2-5: In vitro transcripts obtained 
 using a template pFO108A DNA digested with 
 EcoRI and a second restriction enzyme: 
 lane 2: HindIII; lane 3: PstI; lane 4: 
 XbaI; lane 5: HincII. The positions of 
 the unextended and the extended primers 
 are indicated by arrows.

Indeed occurred in cases where no specific transcript was observed in 1.5% 
 agarose/formaldehyde gels. As a control for these experiments HeLa cell 
 polysomal RNA was used as a template in the primer extension reaction. The 
 results, presented in Fig. 8, lane 1, show that primers extended after 
 hybridization to this control RNA appear as a doublet, indicating that more 
 than one of the multiple HeLa H4 histone mRNAs (20) shares enough sequence 
 homology with the 64 bp primer fragment from pFO108A to form stable hybrids; 
 differences in the 5' leader probably account for the microheterogeneity 
 observed in the extended primer molecules. Lanes 2-5 of Figure 8 show the 
 extended primers obtained with RNA transcribed in vitro, using as a template 
 pFO108A DNA digested with EcoRI plus either HindIII (lane 2), PstI (lane 3), 
 XbaI (lane 4) or HincII (lane 5). In all cases, a single band, comigrating 
 with one of the two extended primers obtained with HeLa polysomal RNA, was 
 observed.
These results confirmed the previous finding that specific and accurate initiation of transcription of the H4 histone gene in pF0108A does occur in vitro. Most interesting, specific initiation of transcription was observed with all of the templates analyzed, although in some cases (pF0108A DNA digested with restriction endonucleases XbaI or HincII), no specific, α-amanitin sensitive run-off transcript was detected.

Taken together, these results suggest that truncating the pF0108A template inside the coding sequence or within 800 nucleotides of the 3' end of the H4 gene has an effect on the production of run-off transcripts of the expected size. However, initiation of transcription still occurs. Further examination is clearly required to assess the mechanism by which in vitro transcription is affected by DNA sequences located downstream from the site of transcription initiation, as well as the possible specificity of the DNA sequences involved.

During the course of these studies it was observed that as the template was truncated closer to the 3' end of the H4 histone gene, thus making the H4 transcript shorter, an α-amanitin insensitive transcript became larger (open circles in Fig. 7). The initiation site for this transcript was determined to be close to the Smal site that is furthest from the H4 gene, with its transcription occurring in the opposite orientation (white arrow in Fig. 1). We have previously shown that the smaller EcoRI/HindIII fragment of the pF0108 insert contains at least one member of the Alu family of repetitive DNA sequences (45 and unpublished results). Since it has been reported that at least some members of the Alu family are transcribed in vitro as well as in vivo (46-49), we conclude that this α-amanitin insensitive transcript (Fig. 7) is the product of the Alu family DNA sequence present in pF0108. In vitro transcription experiments performed using pF0108A as a template show a pattern of transcription similar to that described for pF0108, except that no Alu transcript is produced.

DISCUSSION

We have presented sequencing data supporting the presence of a seemingly functional human H4 histone gene in clone XHHG 41 (15). This gene has the capacity to code for a typical H4 histone protein, and contains several putative regulatory sequences as well as several histone gene-related motifs in its flanking regions. Previous results had suggested that this H4 histone gene is completely homologous to one of the major subspecies of H4 mRNA found in HeLa cells, as shown by its ability to protect that specific H4 mRNA.
To further test the functionality of this H4 histone gene, we have optimized an in vitro system capable of supporting specific initiation of transcription of the H4 gene. The products obtained have been analyzed in two ways: direct sizing of the transcripts and primer extension analysis (39). Transcription of the H4 gene in vitro was found to be dependent on RNA polymerase II activity and appears to initiate at the correct in vivo initiation site.

Several genes transcribed by RNA polymerase II share certain consensus sequences, located upstream from the 5' end of the genes, which are suspected of having some regulatory function (22). Analysis of the DNA sequences preceding the H4 histone gene present in pF0108A indicates the presence of several of these conserved regions (Fig. 2). To test the effect of putative regulatory regions located upstream from the 5' end of the human H4 histone gene, a series of deletions encompassing this region of the DNA was constructed. Analysis of the in vitro transcripts obtained indicated that in this case, as previously described for other genes, no sequences upstream from the TATA box are required for the in vitro transcription of the human H4 histone gene by RNA polymerase II. It is anticipated that the sequences required for transcription in vivo may differ considerably from the in vitro requirements as has been observed in most other systems studied (28).

We are currently assessing the sequences necessary to support transcription of human histone genes in vivo.

An interesting observation made during the course of these studies concerns the involvement of sequences downstream from the 3' end of the gene in the production and/or stability of in vitro transcripts. Sequences located up to 800 bp downstream from the 3' end of the pF0108 H4 histone gene seem to have an effect on the production of a run-off transcript of defined length. Since initiation of transcription is not affected (Fig. 8), it is most likely that this effect is at the level of elongation and/or stability of the in vitro transcripts. These possibilities can be tested using the in vitro transcription system described.

ACKNOWLEDGEMENT

These studies were supported by grants from the March of Dimes - Birth Defects Foundation (1-813), the National Science Foundation (PCM 80-18075 and PCM 81-18951) and the National Institutes of Health (GM 32010).
REFERENCES