Volume 10 Number 21 1982 Nucleic Acids Research

Chicken histone H5 mRNA: the polyadenylated RNA lacks the conserved histone 3' terminator sequence

Paul A.Krieg, Allan J.Robins, Alan Colman and Julian R.E.Wells

Department of Biochemistry, University of Adelaide, South Australia 5001, Australia

Received 30 July 1982; Accepted 28 September 1982

ABSTRACT

Using 3 overlapping cDNA clones we have determined the nucleotide sequence of chicken histone H5 mRNA. The mRNA does not contain the 23 base conserved sequence element[1] that is present at the 3' end of cell-cycle regulated histone mRNAs. Although the RNA is polyadenylated it lacks the 3' AAUAAA sequence.

INTRODUCTION

Histones H1 and H5 share common structural features[2] and occupy similar positions, outside the nucleosome, in chromatin structure[3]. Observed similarities between the primary sequences of the H1 and H5 proteins have led to the suggestion that the genes coding for H1 and H5 may be related[4].

However, a number of observations indicate that H1 and H5 may not be as closely related as the protein sequence data would imply. Unlike the majority of the histone mRNAs, which lack a poly-A tail[5], H5 mRNA has been shown to be polyadenylated[6], and while H5 may be synthesized in the non-dividing reticulocyte[7], the expression of the major classes of histones, including H1, appears to be closely coupled to DNA synthesis[8,9].

In this paper we present the nucleotide sequence of H5 mRNA. The determination of the structure of H5 mRNA and the analysis of features of the nucleotide sequence should help to clarify the relationship of H5 to the other cell-cycle regulated histone genes.

MATERIALS AND METHODS

Sizing of H5 mRNA. Total chick embryo RNA (20 µg), prepared by the guanidine-HCl extraction procedure[10], was electrophoresed in the presence of 10 mM methyl mercury hydroxide in 1.5% agarose gels, as described by Bailey and Davidson[11]. The gels were treated with mercaptoethanol, stained

0305-1048/82/1021-6777$2.00/0

with ethidium bromide, photographed and the RNA transferred to aminophenylthioether (APT) paper[12]. Hybridisation of filter-bound material with the ^{32}P-labelled insert of the H5 cDNA clone, pCH5-01[13], labelled by nick-translation[14], was performed at 42°C as described[15], except that 1% glycine was included in the prehybridisation buffer. After washing[12], filters were exposed to X-ray film at -80°C in the presence of an intensifying screen.

Isolation of H5 cDNA clones. The library of cDNA clones used in these experiments were prepared as described previously[13]. Identification of further H5 recombinants was achieved using the colony-screening procedure[16], and ^{32}P-labelled insert isolated from the H5 recombinant pCH5-01[13].

Restriction enzyme digestion of cDNA clones. All digestions were carried out using restriction endonucleases obtained from New England Biolabs, and under the digestion conditions recommended by the manufacturer.

DNA sequence analysis. Restriction fragments of the cDNA clones were subcloned into the M13 mp83 vector[17] and DNA sequencing was carried out by the chain-termination method of Sanger et al.[18]. Wherever possible, fragments were sequenced in both orientations.

RESULTS

Sizing of chicken H5 mRNA. The insert of the chicken H5 cDNA clone, pCH5-01[13], was labelled by nick-translation and used in association with the Northern transfer procedure to determine the size of mature H5 mRNA. Total chicken embryo RNA was electophoresed on an agarose gel, transferred to APT-paper and probed with the labelled H5 sequences. The data in Fig. 1 shows that the H5 probe detected a single RNA species with a length of 900 ± 50 bases. This estimate is somewhat less than the 1100 bases obtained[6], using polyacylamide gels run in formamide[6,20].

H5 protein contains 189 amino acids[19], and so 567 bases of the 900 bases of H5 mRNA are required for protein coding. The 5' and 3' non-coding regions of H5 mRNA will therefore have a combined length of approximately 300 bases.

Characterisation of recombinants containing H5 mRNA sequences. The H5 cDNA clone initially isolated, pCH5-01[13], was only 350 bp long and contained a portion of the 5' end of the mRNA sequence. In order to identify further recombinants containing H5 mRNA sequences, a library of cDNA clones prepared from 11-18S chicken reticulocyte polyadenylated RNA was screened

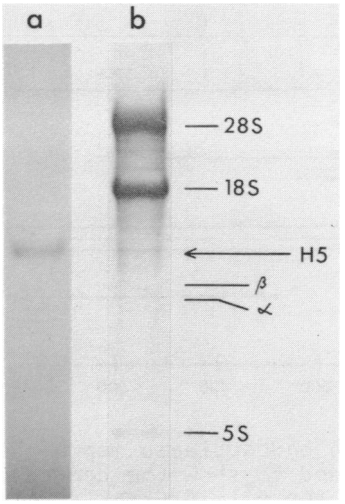

Figure 1: Sizing of H5 mRNA. Total 6-day chicken embryonic RNA was electrophoresed on a 1.5% agarose gel containing 10 mM methyl mercury hydroxide. After transfer to APT paper the RNA was probed with H5 sequences as described in Materials and Methods. a) Autoradiograph of RNA probed with H5 sequences. b) The same track visualized by ethidium bromide staining before transfer. The positions of 28S, 18S and 5S ribosomal RNA and of the adult chicken α and β globin mRNAs, (detected by autoradiography), are indicated.

using nick-translated insert of pCH5-01. 1200 Ampicillin-sensitive colonies were screened by the colony-hybridisation method of Grunstein and Wallis[16], and 9 positively-reacting clones were detected. As the cDNA library had been prepared from mRNA estimated to be 4-fold enriched for H5 mRNA, the number of H5 clones isolated suggests that H5 mRNA represents about 0.2% of chicken reticulocyte polyadenylated RNA.

When the Pst-excisable inserts of the 9 H5 cDNA clones were sized, three were found to contain a total length of inserted DNA of greater than 500 base pairs. These recombinants, named pCH5-02, pCH5-03 and pCH5-04 carried insertions of about 650, 850 and 550 base pairs respectively. Since these total fragment lengths included a poly C/G segment, (averaging approximately 20-30 base pairs), at each end of the inserted sequence, it was clear that not even the largest clone, pCH5-03, would carry the entire 900 base nucleotide sequence of H5 mRNA.

Cleavage of the three longest clones with 6-base specificity restriction endonucleases showed that the clones overlapped to a large

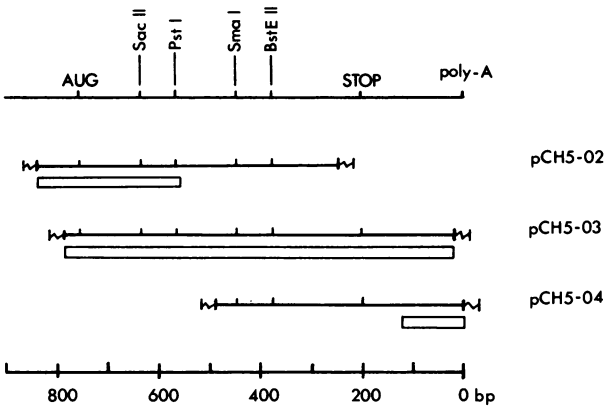

Figure 2: Restriction endonuclease map of H5 cDNA clones. The clones have been aligned to show the degree of overlap. The symbol ⋈ indicates the poly G/C tails and the open-bars below the clones show the regions for which the DNA sequence was determined.

extent, (Figure 2). Although pCH5-03 was the longest clone, the restriction mapping showed that it did not extend as far in the 5' direction as pCH5-02, or as far in the 3' direction as pCH5-04. (DNA-sequencing later revealed that only pCH5-04 contained the poly-A sequence and 3' terminal nucleotides of H5 mRNA). Together, the total length of mRNA sequence delineated by the clones was about 850 bases, sufficient to account for the great majority of the H5 mRNA sequence.

The nucleotide sequence of chicken H5 mRNA. The complete nucleotide sequence derived for the H5 cDNA clones is presented in Figure 3. The total sequence is 848 bases long, comprising 80 bases of 5' untranslated region, 570 bases of protein coding region, (including the AUG initiation codon, which is not expressed in the mature H5 protein), and 198 bases of 3' non-coding sequence, to the beginning of the poly-A tail. The sequence of the coding region predicts precisely the known amino acid sequence of the arginine, (position 15), variant of chicken histone H5[19]. The original H5 cDNA clone isolated, pCH5-01, encoded the glutamine variant of H5, but, of the 180 bases that have been sequenced in both the glutamine and arginine coding cDNA clones, the only difference observed is the single base change needed to bring about the glutamine⟶arginine substitution.

The sequence of the 5' non-coding segment of H5 mRNA coincides with the sequence reported by Ruiz-Vazquez and Ruiz-Carillo[20], except for a single base change, C⟶T, at position -25. Their sequence contains a

```
                                                    5' GCCCCACATCCGTTGT

TGCTGGCGGCTCCTTTTTTTAAGCTCCCTAACCCCAGTGCCCTGCCGTGGGGTGAAGCGGCGGCC    ATG

thr glu ser leu val leu ser pro ala pro ala lys pro lys arg val lys ala
ACG GAG AGC CTG GTC CTA TCC CCA GCC CCA GCC AAG CCC AAG CGG GTG AAG GCA   18

ser arg arg ser ala ser his pro thr tyr ser glu met ile ala ala ala ile
TCG CGG CGC TCG GCA TCG CAC CCC ACC TAC TCG GAG ATG ATC GCG GCG GCC ATC   36

arg ala glu lys ser arg gly gly ser ser arg gln ser ile gln lys tyr ile
CGT GCG GAA AAG AGC CGC GGC GGC TCC TCG CGG CAG TCC ATC CAG AAG TAC ATC   54

lys ser his tyr lys val gly his asn ala asp leu gln ile lys leu ser ile
AAG AGC CAC TAC AAG GTG GGC CAC AAC GCC GAT CTG CAG ATC AAG CTC TCC ATC   72

arg arg leu leu ala ala gly val leu lys gln thr lys gly val gly ala ser
CGA CGT CTC CTG GCT GCC GGC GTC CTC AAG CAG ACC AAA GGG GTC GGG GCC TCC   90

gly ser phe arg leu ala lys ser asp lys ala lys arg ser pro gly lys lys
GGC TCC TTC CGC TTG GCC AAG AGC GAC AAG GCC AAG AGG TCC CCC GGG AAG AAG  108

lys lys ala val arg arg ser thr ser pro lys lys ala ala arg pro arg lys
AAG AAG GCC GTC AGG AGG TCC ACG TCT CCC AAG AAG GCG GCG AGG CCC AGG AAG  126

ala arg ser pro ala lys lys pro lys ala thr ala arg lys ala arg lys lys
GCC AGG TCA CCG GCC AAG AAG CCC AAA GCC ACC GCC AGG AAG GCC AGG AAG AAG  144

ser arg ala ser pro lys lys ala lys lys pro lys thr val lys ala lys ser
TCG CGG GCA AGC CCC AAG AAG GCC AAG AAG CCA AAG ACT GTT AAG GCC AAG TCG  162

arg lys ala ser lys ala lys lys val lys arg ser lys pro arg ala lys ser
CGG AAG GCC TCC AAG GCC AAG AAG GTG AAG CGG TCG AAA CCC AGA GCC AAG TCT  180

gly ala arg lys ser pro lys lys lys
GGC GCC CGG AAA TCG CCC AAG AAG AAG TGA     GCAGCCTGGGGGCTTTGTCCAGGCTCTC

CCCATTGGTTTCTGTAAATAGCTTTTGCCTTTATTTTTACCTCTTTCTATTTGCAAATTTTATAAGTTGATC

TATTCCTAAGAGCTAAAACAAGGCAACGAATGAAAGAAAAAAAAGAAACAAAAATGGAACTTCTTCCATATGG

AAGAGTTCCCGTTTATAAAAGCTAAAAAAAA polyC 3'
```

Figure 3: Nucleotide sequence of H5 cDNA. The complete sequence was derived from the three H5 cDNA clones illustrated in Figure 2. Also shown is the predicted amino acid sequence which is in complete agreement with the H5 protein sequence determined by Briand et al.[19]. The arrows below the sequence indicate the residues that may be involved in secondary structure. (Numbered from +1 AUG).

total of 115 bases in the 5' region, and analysis of the chicken H5 genomal clone, (manuscript in preparation), suggests that this probably represents the complete 5' sequence of H5 mRNA. The appearance of inverted repeats in the 5' region and the possibility of sequences shared with H1 genes have already been discussed in detail[20].

The most interesting aspects of the H5 mRNA sequence are found in the 3' untranslated region. The 3' sequence is extremely A-T rich, so that, while the 5' and protein-coding regions of the mRNA have a G-C content of 65 % and 63 % respectively, the G-C content of the 3' untranslated region is only 36 %. Particularly conspicuous is the sequence between residues +706 and +726, in which 18 of the 21 bases are adenine.

Ribonuclease digestion studies[6] and the side-by-side comparison of H5 genomal sequences (in preparation) and 3' cDNA sequences (Fig. 3) show that H5 mRNA is polyadenylated. Divergence between genomal and cDNA sequences occurs precisely at the junction of the T residue and the eight A residues (attached to poly C) at the 3' end of the cDNA clone (Fig. 3). All other oligo-A runs in the 3' region correspond to genomal DNA sequences. However, the 'AAUAAA' sequence motif, usually found near the 3' terminus of poly-adenylated mRNAs[21] is absent from the H5 mRNA sequence. The variant form, AUUAAA, observed in angler-fish somatostatin mRNA[22] and rat amylase mRNA[23] also does not appear in the H5 mRNA.

Perhaps the most important observation concerning the H5 mRNA sequence is that the 3' end of the mRNA does <u>not</u> contain the 23 base conserved homology element, (consensus 5' AACGGC$^{C}_{U}$CUUUUCAG$^{G}_{A}$GCCACCA 3')[1], that is found at the 3' end of all other reported histone mRNA sequences, (review ref. 24). The conserved homology block contains a 16 base hyphenated inverted repeat that is postulated to form a hairpin-loop structure, immediately adjacent to the 3' terminus of the histone mRNA[24]. Although lacking the histone-specific inverted repeat sequence, the H5 mRNA does contain sequences that have the potential to form secondary structures, and the two most stable of these are illustrated in Figure 4. The first structure, (Fig. 4A), involves nucleotides +576 to +598 and is situated just 2 bases downstream from the UGA stop codon. It consists of an 8 bp stem with a 7 base loop and the ΔG value for the arrangement is calculated at -50 KJ/mole[25]. The second hairpin loop structure, (Figure 4B), includes residues +724 to +759 and is thus situated immediately adjacent to the 3' terminus of the H5 mRNA. The ΔG value of this secondary structure is -52.8 KJ/mole. By way of comparison, the 6 base pair hairpin loop at the

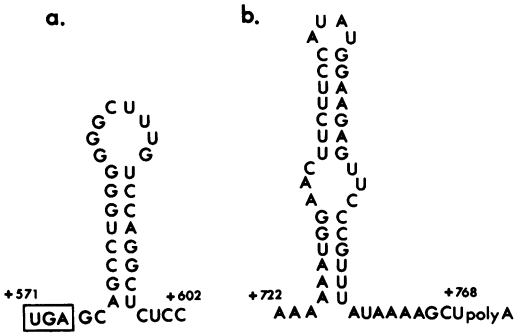

Figure 4: Potential secondary structure in the 3' non-coding region of H5 mRNA. The hairpin loop structure in 4b is the most thermodynamically stable of several optional structures in this region.

end of the normal histone mRNAs has a ΔG value of up to -35.3 KJ/mole, although the stability may be considerably less, depending on the exact sequence composing the hairpin.

DISCUSSION

The determination of the nucleotide sequence of chicken H5 mRNA allows a number of comparisons to be made between the structure of H5 mRNA and of the other histone mRNAs.

Using Northern analysis, (Figure 1), we have shown that the H5 mRNA is approximately 900 bases long. The DNA sequence indicates that, of this total length, 198 bases are in the 3' non-coding region of the mRNA. This A-T rich 3' sequence is considerably longer than the equivalent region of the other chicken histone mRNAs which have a length of about 50 to 70 bases, (unpublished data). H5 mRNA is polyadenylated[6], but the mRNA does not contain the sequence, 5' AAUAAA 3', which is very commonly observed towards the 3' end of poly-A containing mRNAs, and which has been demonstrated to be important for the polyadenylation of SV40 late gene mRNAs[26]. While H5 mRNA does contain a related sequence, UAUAAA, just four bases from the poly A tail, this would appear to be unusually close to the end of the mRNA to be the true polyadenylation signal, and this sequence may be present at this position due to chance, especially since the 3' end of the mRNA is very A-T rich. The AAUAAA sequence has been observed as little as 9 bases from the poly-A tail (for example encephalomyocarditis virus mRNA[27]), but, in general it is found between 11 and 30 bases from the

3' terminus of the mRNA[26]. We may conclude then that H5 mRNA resembles some viral RNAs, (for example, the Foot and Mouth Disease Viruses[27]) and some plant mRNAs, (small subunit of ribulose-1,5-biphosphate carboxylase[28]), in lacking an obvious polyadenylation signal.

Undoubtedly the most conspicuous feature of histone mRNAs is the 23 base conserved homology block[1], located just before the 3' end of the mature histone mRNA, and which has been shown to be necessary for the correct termination of histone mRNA transcription[29]. This conserved sequence element has been detected, virtually unchanged, at the 3' end of the histone genes of a wide range of species[24], and is present in all chicken histone genes, including those coding for H1, (unpublished data). Chicken H5 mRNA does not contain the histone 23 base conserved sequence and inspection of the 3' segment of H5 mRNA does not reveal any regions showing obvious residual homology with this sequence. The absence of the homology block from H5 mRNA suggests that, whatever the specific functional importance of the 23 base conserved sequence to the correct transcription or processing of the histone mRNAs, H5 mRNA transcription does not share this requirement.

While H5 mRNA lacks the histone inverted repeat sequence, the 3' region of H5 mRNA is not devoid of secondary structure. The nucleotide sequence suggests the presence of two stable hairpin-loop structures, (Figure 4), one of which is located extremely close to the 3' terminus of the mRNA. The large free energy values for these hairpin loops suggest that, in the absence of other constraints, these features will be present in the mature mRNA and that therefore H5 mRNA, like prokaryote mRNA[30], eukaryote polymerase III transcripts[31], and the cell-cycle regulated histone mRNAs[24], terminates in an RNA hairpin structure.

In summary, the examination of the nucleotide sequence has shown that chicken H5 mRNA does not have the structure expected for a normal histone mRNA. Since it lacks the 23 base conserved sequence that has been shown to be essential for the correct termination of the other histone mRNAs[29], we may expect that the H5 mRNA transcript is processed by an independent mechanism. Therefore, despite the structural similarities observed between the H1 and H5 proteins[2,4], the genes coding for H1 and H5 do not appear to be closely related. Analysis of the H5 genomal gene should provide more information on the relationship of H5 to the other cell-cycle regulated histones.

REFERENCES

1. Busslinger, M., Portmann, R. and Birnstiel, M.L. (1979). Nuc. Acids Res. 6, 2997-3008.
2. Aviles, F.J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C. and Bradbury, E.M. (1979). Eur. J. Biochem. 88, 363-71.
3. Johns, E.W. (1971) in Histones and Nucleohistones (Ed. D.M.P. Phillips), Plenum Press, London.
4. Yaguchi, M., Roy, C., Dove, M. and Seligy, V. (1977). Biochem. Biophys. Res. Commun. 76, 100-106.
5. Kedes, L.H. (1979). Ann. Rev. Biochem. 48, 837-870.
6. Molgaard, H.V., Perucho, M. and Ruiz-Carrillo, A. (1980). Nature 283, 502-504.
7. Appels, R. and Wells, J.R.E. (1972). J. Molec. Biol. 70, 425-434.
8. Rickles, R., Marashi, F., Sierra, F., Clark, S., Wells, J.R.E., Stein, J. and Stein, G. (1982). Proc. Natl. Acad. Sci. USA 79, 749-753.
9. Wu, R.S. and Bonner, W.M. (1981). Cell 27, 321-330.
10. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979). Biochemistry 24, 5294-5299.
11. Bailey, J.M. and Davidson, N. (1976). Anal. Biochem. 70, 75-85.
12. Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 5350-5353.
13. Krieg, P.A., Robins, A.J., Gait, M.J., Titmas, R.C. and Wells, J.R.E. (1982). Nucl. Acids Res. 10, 1495-1502.
14. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). J. Molec. Biol. 113, 237-251.
15. Thomas, P.S. (1980). Proc. Natl. Acad. Sci. U.S.A. 77, 5201-5205.
16. Grunstein, M. and Wallis, J. (1979). Meth. Enzm. 68, 379-389.
17. Messing et al. (in press).
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
19. Briand, G., Kniecik, D., Sautiere, P., Wouters, D., Borie-Loy, O., Biserte, G., Mazen, A. and Champagne, M. (1980). FEBS Lett. 112, 147-151.
20. Ruiz-Vazquez, R. and Ruiz-Carrillo, A. (1982). Nucleic Acids Res. 10, 2093-2108.
21. Proudfoot, N.J. and Brownlee, G.G. (1976). Nature 263, 211-214.
22. Hobart, P., Crawford, R., Shen, L.P., Pictet, R. and Rutter, W.J. (1980). Nature 288, 137-141.
23. MacDonald, R.J., Crerar, M., Swain, W., Pictet, R., Thomas, G. and Rutter, W.J. (1980). Nature 287, 117-122.
24. Hentschel, C.C. and Birnsteil, M.L. (1981). Cell 25, 301-313.
25. Riesner, personal communication.
26. Fitzgerald, M. and Shenk, T. (1981). Cell 24, 251-260.
27. Porter, A.G., Fellner, P., Black, D.N., Rowlands, D.J., Harris, T.J.R. and Brown, F. (1978). Nature 276, 298-301.
28. Bedbrook, J.R., Smith, S.M. and Ellis, R.J. (1980). Nature 287, 692-697.
29. Birchmeier, C., Grosschedl, R. and Birnsteil, M.L. (1982). Cell 28, 739-745.
30. Pribnow, D. (1979) in Biological Regulation and Development, 1, R.F. Goldberger, ed. (New York, Plenum Press) pp. 250-277.
31. Korn, L. and Brown, D. (1978). Cell 15, 1145-1156.