The chicken H5 gene is unlinked to core and H1 histone genes

Paul A.Krieg, Allan J.Robins, Richard D'Andrea and Julian R.E.Wells

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

Received 15 December 1982; Accepted 11 January 1983

ABSTRACT

An H5 cDNA clone was used to select H5 genomic recombinants from a chicken Charon 4A library. DNA sequence analysis shows that the H5 gene contains no introns. Putative 5' promoter elements and a 3' polyadenylation site are present within the 1.8 kb of DNA examined. Analysis of 41 kb of DNA surrounding the H5 gene shows that it is not closely linked to either H1 or core histone genes.

INTRODUCTION

The histone H5 is tissue-specific, being found in the erythroid cells of birds, fish, reptiles and amphibia¹. H5 shows structural² and sequence³ similarities to histone H1, and occupies an equivalent position in the chromatin⁴. Accumulation of H5 is correlated with a reduction in genetic expression and cell proliferation⁵, and therefore its biological properties closely resemble those of the mammalian histone H1⁶. With the exception of the sea urchin orphon genes⁷ and possibly an unusual H3 gene in chickens⁸, the histone genes exist in clusters in genomic DNA, either in simple tandem repeats or in other less-ordered arrangements⁹. We have now determined the DNA-sequence of the chicken H5 gene and studied its chromosomal organisation in relation to the core and H1 histone genes. Isolation of cDNA clones containing histone H5 sequences¹⁰, has permitted the selection of recombinants containing the chromosomal H5 gene from a genomic library. Our results indicate that the H5 gene is unique, that it contains no introns and that, as noted previously¹¹, the polyadenylated H5 mRNA lacks the terminator element typical of other histone messengers. Specific probes for H2A, H2B, H3, H4 and H1 histone genes failed to hybridise to a 41 kb region of genomic DNA containing the H5 gene. We therefore conclude that in the chicken genome the H5 gene is not linked to the other histone genes.

MATERIALS AND METHODS

Isolation of Genomal Clones: Approximately 5×10^5 recombinant Charon 4A phage were screened for clones containing H5 sequences by the in situ plaque hybridisation procedure of Benton and Davis¹². The hybridisation probe was a cDNA clone¹⁰ containing 5' untranslated and coding sequences of H5 mRNA, and labelled by nick-translation¹³. Recombinant phage overlapping the longest H5 containing clone (λ CH5-01, see Fig. 1) were selected from the genomal library in similar fashion, using as probes, ³²P-labelled fragments representing the extreme 5' and 3' ends of the insert of λ CH5-01 (1.5 kb EcoRI/HindIII fragment and 0.9 kb BamHI/EcoRI fragments respectively).

Mapping and Analysis of Genomal Clones: All restriction enzyme analyses were carried out under conditions recommended by the suppliers. Filter hybridisations were by the method of Southern¹⁴ as modified by Wahl et al.¹⁵.

DNA Sequence Analysis: Restriction fragments from insert DNA were subcloned into the M13mp83 vector and DNA sequencing carried out by the chain-termination method of Sanger et al.¹⁶.

RESULTS AND DISCUSSION

Isolation and Characterisation of H5 Genomal Clones: An H5 clone¹⁰ was used to screen a chicken recombinant DNA library. Of the recombinants isolated, the longest clone, λ CH5-01, was selected for detailed characterisation (Figure 1). Hybridisation with H5 probe and comparison of the restriction patterns of the cDNA¹¹ and genomal clones allowed the position of the H5 coding sequences and the direction of transcription of the H5 gene to be determined. The H5 gene is located almost centrally within the 16.6 kb insert of λ CH5-01.

The sequence of about 1.8 kb of DNA surrounding the H5 gene has been determined using M13 cloning and the dideoxy chain termination sequencing techniques (Figure 2). The sequences comprise 893 bases 5' to the ATG initiator codon, 573 coding bases (including the initiation and termination codons), and 377 bases 3' to the termination codon. Inspection of the 5' flanking sequences of the gene shows that the H5 gene has the potential promoter sequence 5'-CTTAAAT-3', centred 138 bases from the initiation codon. Examination of sequences to -893 reveal no other candidates for a more highly-conserved TATA box. While the H5 promoter differs considerably from the canonical sequence 5'-TATA^{TAT}_{ATA}-3'¹⁷, it resembles the promoter sequences 5'-GTAAAA-3' and 5'-TTAAAA-3', which also lack the 'ATA'

Figure 1. Restriction map of chicken genomic DNA surrounding the histone H5 gene. The segment of DNA is contained within the recombinant λ CH5-01. The blocked-in region represents the protein-coding sequences of the H5 gene and the direction of transcription is from left to right, as indicated by the arrow. Eight clones containing H5 sequences were isolated but only the longest, λ CH5-01, has been mapped extensively. The λ CH5-01 insert contains no EcoRI sites, and the EcoRI sites at the end of the clone are artificial having been created during construction of the genomic library. The region of the clone for which the DNA sequence has been determined (Figure 2) is indicated by the horizontal bracket.

sequence and which are detected upstream from the chicken lysozyme¹⁸ and *Dictyostelium* actin¹⁹ genes respectively. Using a technique that clones the entire 5' untranslated sequences of mRNA with high efficiency²⁰, Ruiz-Vasquez and Ruiz-Carrillo²¹ have reported the DNA-sequence of an H5 recombinant that terminates at -109 on our genomic DNA sequence. Since there is agreement between the 5' sequence of this cDNA and the genomic sequence reported here (except for an insert GGC GGC AGC present for the cDNA at a position corresponding to -91 in the genomic sequence, and a T for a C at -25), and because the cDNA sequence ends at a position 30 bases downstream from the TATA box, we tentatively locate the capping site of the H5 gene at the C-residue at -109 in the genomic sequence. The 5' region of the H5 gene does not contain any sequences that show close homology to the 'CAAT' sequence which is observed 30 to 60 bases upstream from the TATA box in most eukaryote genes²².

At the 3' end of the H5 gene, the site of polyadenylation may be determined as the point of divergence of the 3' sequences of the cDNA and genomic clones, and corresponds to nucleotide +768 of the H5 genomic sequence (arrowed in Figure 2 and discussed in detail in reference 11). The distance from the putative cap site (-109) to the poly-A addition site is some 880 bases, in good agreement with the estimate of 900 bases for mature

```

-893                               -850
CTGGAGGCATGGCTCTCATCTCCAGATGAGAGCTCTGCGTGTCCCAAGCCTCTGGTGGGGA

-800                               -750
GCAAAAGGGACAGGGCGGAGTGGAGACAGGCTTATTGGACCATGGGGTTCAGCTGCCAAACCTATTTTTAGCCTTTTGGCA

-700
GAAAAAAGTGCACCACCACCCCTCCCATACATACCACCCCTGGTAAGTGGGAGCCACCACAAAACCCAGAGTGCCTAGGGCA

-650                               -600
GACAGAAGAGTGGAGACCCCGGTGTCTGTCCAATGTCCCCCCTCATTATTCCTACATTCTCCCATGTCTCCTTGTCT

-550                               -500
CCATCCAGTGTTCCTAACACCCCTTCTGAGTCTCCTTGCCACGTCTCCCTGCTCCTCCTGTCCCCAGCAGAGCTCCTCCT

-450
TGCCCTGCACATTACCTCCTTGTCTCTGTATCTCCCTCCAGCACCTATCCTTGTCTACCGAGTGTCAAGTCAAAAGTACTCC

-400                               -350
TTTTCCCATCACATCTCTCCTGTCCCTGTGTCACACCCAGTGTCCACCCCTGTCCCCACCATGTCCATCCCATCCAC

-300                               -250
CGTGTCTCACCTATCCCAACCGTATCCCTCTCCATCCCTTAATGTCCAACCCAGCGTCCGCCCTTTCCCATCACATCC

-200
TTCTGGTCCCAACCTGTCTCCTCCTGCTGCCACGGCATGTCTCGGGGGCGGGGAGAGGGGGACACGGGGACAGGCAG

-150                               *                               -100
TCCTCCCCGCGGTCCGTGCCACCCCTAAATGCGTGTGTGGCGACGCGGGCGCAGACGACGCGGCGGCGGACGAGGA

-50
GCAGCCCCACATCGTTGTTGTCTGCGGGCTCTTTTTTAAGCTCCCTAACCCAGTGCCTGCGGTGGGGTGAAGCGCGGCG

1                               50
thr glu ser leu val leu ser pro ala pro ala lys pro lys arg val lys ala ser arg
ATG ACG GAG AGC CTG GTC CTA TCC CCA GCC CCA GCC AAG CCC AAG CGG GTG AAG GCA TCG CGG

100
arg ser ala ser his pro thr tyr ser glu met ile ala ala ile arg ala glu lys ser
CGC TCG GCA TCG CAC CCC ACC TAC TCG GAG ATG ATC GCG GCG GCC ATC CGT GCG GAA AAG AGC

150
arg gly gly ser ser arg gln ser ile gln lys tyr ile lys ser his tyr lys val gly his
CGC GGC GGC TCC TCG CGG CAG TCC ATC CAG AAG TAC ATC AAG AGC CAC TAC AAG GTG GGC CAC

200                               250
asn ala asp leu gln ile lys leu ser ile arg arg leu leu ala ala gly val leu lys gln
AAC GCC GAT CTG CAG ATC AAG CTC TCC ATC CGA CGT CTC CTG GCT GCC GGC GTC CTC AAG CAG

300
thr lys gly val gly ala ser gly ser phe arg leu ala lys ser asp lys ala lys arg ser
ACC AAA GGG GTC GGG GCC TCC GGC TCC TTC CGC TTG GCC AAG AGC GAC AAG GCC AAG AGG TCC

350
pro gly lys lys lys lys ala val arg arg ser thr ser pro lys lys ala ala arg pro arg
CCC GGG AAG AAG AAG AAG GCC GTC AGG AGG TCC ACG TCT CCC AAG AAG GCA GCG AGG CCC AGG
G

400
lys ala arg ser pro ala lys lys pro lys ala thr ala arg lys ala arg lys lys ser arg
AAG GCC AGG TCA CCG GCC AAG AAG CCC AAA GCC ACC GCC AGG AAG GCC AGG AAG AAG TCG CGG

450                               500
ala ser pro lys lys ala lys lys pro lys thr val lys ala lys ser arg lys ala ser lys
GCA AGC CCC AAG AAG GCC AAG AAG CCA AAG ACT GTT AAG GCC AAG TCG CGG AAG GCC TCC AAG

550
ala lys lys val lys arg ser lys pro arg ala lys ser gly ala arg lys ser pro lys lys
GCC AAG AAG GTG AAG CGG TCG AAA CCC AGA GCC AAG TCT GGC GCC CGG AAA TCG CCC AAG AAG

lys
AAG TGA GCAGCCCGGGCTTTGCCAGGCTCTCCCATTTGTTCTGTAAATAGCTTTTGCTTTATTTTACCTCTT
T A G T

650                               700
TCTATTTCGAAATTTTATAAGTTGATCTATTCTTAAGAGCTAAACAAGGCAACGAATGAAGAAAAAAGAAACAAAATGG

750                               800
AACTTCTTCATATGGAAGAGTTCCCGTTTATAAAAGCA + ACTTCTGAGTGTTTATTCATCTTGCCGTGGGTCTGGAAGG
T

850
TTGCGTGGGCGGTGGAGAGCTTGGGGGAGGTAGAGGGCTGTGAGAGCAGGAGGAGGAGGGGACTCCTCCTGTCCA

900                               950
TAGGAGTGAGGCACAGCCCGCCAGCCAGCGGTGAGGGCTGGAGATAACAGTGGCG

```

H5 mRNA sized on methyl-mercury-hydroxide gels (in which a poly-A tract would also be present)¹¹. The nucleotide sequence of the genomic H5 gene is colinear with the sequence derived from H5 cDNA clones, indicating that the H5 gene contains no introns. This observation explains the failure to detect precursor forms of H5 mRNA in RNA-blotting experiments^{11,23}.

In general, there is close agreement between the sequences derived from cDNA clones¹¹ and the H5 gene in λ CH5-01, both of which represent the arginine (position 15) allele of the H5 gene. Nucleotide differences exist at 5 positions, as indicated in Figure 2. Within the 573 bases of coding region the genomic sequence agrees with the sequence of the cDNA clones except for a single base change in the third position of the codon for alanine 121 (base 366 in Figure 2 is A, while the cDNA contains a G in this position). The other 4 changes are confined to the 3' untranslated region of the gene. While it is possible that these differences are due to errors accumulated during the cloning procedures, it is also possible that they represent true variation between strains or individuals. Heilig *et al.*²⁴ have reported that sequences of the introns of the single-copy chicken ovalbumin Y gene show up to 1 percent variation between individuals.

Earlier data, obtained with an impure probe, suggested that the H5 gene may be reiterated as much as 10-fold in the chicken genome²⁵, but use of a cloned DNA probe indicates that H5 is more likely to be a single-copy gene. The number of H5 positive plaques obtained from independent screenings of the chicken λ -Charon 4A library (average of one positive per 10^5 plaques) is consistent with this latter value, and in addition, 8 independent isolates of recombinants containing the H5 genomic gene all yielded restriction patterns showing that they overlapped with λ CH5-01 (Figure 4 and unpublished data). However, the most persuasive evidence comes from Southern blotting experiments. If H5 was encoded by several genes one would expect to see multiple hybridising bands on Southern blot

Figure 2. Nucleotide sequence of the chromosomal chicken histone H5 gene and flanking sequences. Bases are numbered from the first base of the initiator codon which is marked as +1, and the coding region is represented as triplets and includes initiator and terminator codons. The DNA sequence presented correctly predicts the amino acid sequence of the arginine (position 15) variant of chicken histone H5²⁸. The proposed TATA box, centred at -138 is underlined, and the putative start of RNA transcription is marked with a star. The poly-A addition site is indicated by a vertical arrow. Bases below the genomic H5 sequence represent differences with sequences derived from cDNA clones of H5 mRNA¹¹. Δ G represents an inserted base in the cDNA sequence.

Figure 3. Southern blot hybridisation analysis of chicken genomic DNA and phage λ CH5-01 DNA with histone H5 probe. (a) Chicken genomic DNA was digested to completion with HindIII or SacI restriction enzymes, fractionated on a 1% agarose gel and transferred to nitrocellulose¹⁴. The filter was hybridised with ³²P-labelled pCH5-03 DNA probe¹¹, which contains the entire protein-coding sequences of the H5 gene, and washed under conditions of low stringency (3 x SSC at 65°C) in order to reveal other bands that cross-reacted with H5 probe, should these exist. (b) Phage λ CH5-01 DNA was similarly digested and probed and washed under high stringency conditions (0.2 x SSC at 65°C). The sizes of the hybridising bands are indicated between the autoradiographs. The positions of the relevant HindIII and SacI sites in the insert of λ CH5-01 are shown.

analysis of genomic DNA when using an H5-specific probe. Digestion of chicken genomic DNA with three enzymes (SacI and HindIII - Figure 3a and EcoRI, unpublished data), each yielded only a single hybridising band when probed with H5 sequences. The EcoRI band has not been accurately sized, but is greater than 20 kb in length, consistent with the data obtained from overlapping genomic clones (Figure 4). The hybridising bands in the SacI and HindIII genomic digestions (6.2 and 9 kilobases, respectively), correlated in size with those observed when λ CH5-01 was similarly digested (Figure 3b). We conclude, therefore, that the gene in λ CH5-01 is representative of the H5 gene in the chicken genomic DNA and that there is only one

Figure 4. Schematic representation of overlapping genomic clones in the histone H5 gene region. Overlapping clones were analysed by restriction mapping and only the recombinants giving the greatest extension in the 5' and 3' direction are presented in the figure.

copy of the H5 gene per haploid genome, although the possibility that there are several H5 genes contained within identical DNA environments cannot be rigorously excluded.

Linkage Analysis of λCH5-01 and Overlapping Clones for H1 and Core Histone Genes: Analysis of genomic recombinants shows that H1 genes are interspersed with core histone genes in the chicken genome²⁶. We have performed experiments to determine whether the H5 gene is similarly interspersed.

Figure 5. Probing of the H5 gene region with histone sequences. Duplicate plaques of λCH-01, λCH5-01, λHE-1, λBE-3 and a control of λCharon 4A vector were transferred to nitrocellulose and hybridised with the gene-specific probes indicated. λCH-01 contains H1 and all four core histone genes²⁶.

Some 50 recombinants, independently isolated from the chicken genomic library and known to contain either H1 or core histone sequences all failed to hybridise to H5-specific probe (data not shown). The recombinant λ CH5-01, containing the H5 gene, similarly failed to hybridise to gene-specific core and H1 histone probes (Figure 5), setting a limit of about 7 kilobases on each side of the H5 gene in which no other histone sequences were present. We have noted intergene regions of up to 9 kb in chicken histone gene clusters and we therefore extended analyses to recombinants overlapping λ CH5-01. These were obtained after re-screening the chicken library with DNA fragments from each end of the λ CH5-01 insert. Together, the overlapping clones that provide the greatest extension represent a continuous fragment of chromosomal DNA of about 41 kilobases in length (Figure 4). Plaques of these three clones, together with a positive control of λ CH-01^{26,27} which contains H1, H2A, H2B, H3 and H4 gene sequences and a negative control of λ -Charon 4A, were grown on a bacterial lawn in duplicate, transferred onto a nitrocellulose grid¹² and hybridised with gene-specific probes (Figure 5). The results show that none of the histone genes present in λ CH-01 hybridises with the extensive region of DNA surrounding the H5 gene. It is possible that other classes of histone genes exist that would not hybridise with the core and H1 genes present in clones such as λ CH-01²⁶. However, the data here show that the well-characterised histone genes are not closely linked to the H5 gene. Further investigations are required to determine the basis for the tissue-specific expression of the H5 gene and the possible biological role of the H5 protein.

REFERENCES

1. Neelin, J.M., Callahan, P.X., Lamb, D.C. & Murray, K. (1964). *Can. J. Biochem.* **42**, 1743-1752.
2. Aviles, F.J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C. & Bradbury, E.M. (1979). *Eur. J. Biochem.* **88**, 363-371.
3. Yaguchi, M., Roy, C., Dove, M. & Seligy, V. (1977). *Biochem. Biophys. Res. Commun.* **76**, 100-106.
4. Torres-Martinez, S. & Ruiz-Carrillo, A. (1982). *Nucleic Acids Res.* **10**, 2323-2335.
5. Weintraub, H., Campbell, G. & Holtzer, H. (1971). *J. Cell. Biol.* **50**, 652-668.
6. Pehrson, J.R. & Cole, R.D. (1981). *Biochemistry* **20**, 2298-2301.
7. Childs, G., Maxson, R., Cohn, R. & Kedes, L. (1981). *Cell* **23**, 651-663.
8. Engel, J.D., Sugarman, B.J. & Dodgson, J.B. (1982). *Nature* **297**, 434-436.
9. Hentschel, C.C. & Birnstiel, M.L. (1981). *Cell* **25**, 301-313.
10. Krieg, P.A., Robins, A.J., Gait, M.J., Titmas, R.C. & Wells, J.R.E. (1982). *Nucleic Acids Res.* **10**, 1495-1502.

11. Krieg, P.A., Robins, A.J., Colman, A. & Wells, J.R.E. (1982). *Nucleic Acids Res.* 10, 6777-6785.
12. Benton, W. & Davis, R. (1977). *Science* 196, 180-182.
13. Rigby, P., Dieckmann, M., Rhodes, C. and Berg, P. (1977). *J. Molec. Biol.* 113, 237-251.
14. Southern, E.M. (1975). *J. Molec. Biol.* 98, 503-517.
15. Wahl, G.M., Stern, M. and Stark, G.R. (1979). *Proc. Natl. Acad. Sci. USA* 76, 3683-3687.
16. Sanger, F., Nicklen, S. & Coulson, A.R. (1977). *Proc. Natl. Acad. Sci. USA* 74, 5462-5467.
17. Cordon, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Keddinger, C. & Chambon, P. (1980). *Science* 209, 1406-1414.
18. Grez, M., Land, H., Giesecke, K., Schutz, G., Jung, A. & Sippel, A. (1981). *Cell* 25, 743-752.
19. McKeown, M. & Firtel, R. (1981). *Cell* 24, 799-807.
20. Land, H., Grez, M., Hauser, H., Lindenmaier, W. & Schutz, G. (1981). *Nucleic Acids Res.* 9, 2251-2266.
21. Ruiz-Vasquez, R. & Ruiz-Carrillo, A. (1982). *Nucleic Acids Res.* 10, 2093-2108.
22. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). *Nucleic Acids Res.* 8, 127-142.
23. Colman, A., Krieg, P.A. & Wells, J.R.E. (1982) (submitted).
24. Heilig, R., Muraskowsky, R., Kloepper, C. & Mandel, J. (1982). *Nucleic Acids Res.* 10, 4363-4396.
25. Scott, A. & Wells, J.R.E. (1976). *Nature* 259, 635-638.
26. Wells, J.R.E. et al. (1982) In: 'Manipulation and Expression of Genes in Eukaryotes', (Nagley, P., Linnane, A., Peacock, W. & Pateman, J. eds.). Academic Press, Sydney (in press).
27. Harvey, R.P., Krieg, P.A., Robins, A.J., Coles, L.S. & Wells, J.R.E. (1981). *Nature* 294, 49-53.
28. Briand, G., Kniecik, D., Sautiere, P., Wouters, D., Borie-Loy, O., Biserte, G., Mazen, A. & Champagne, M. (1980). *FEBS Lett.* 112, 147-151.