Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1936–1940, April 1984
Biochemistry

Evidence for a human histone gene cluster containing H2B and H2A pseudogenes

(gene structure/gene expression/regulatory sequence)

F. MARASHI*, K. PROKOPP†, J. STEIN†, AND G. STEIN*

*Department of Biochemistry and Molecular Biology, and †Department of Immunology and Medical Microbiology, University of Florida College of Medicine, Gainesville, FL 32610

Communicated by George K. Davis, November 7, 1983

ABSTRACT **Not all members of the human histone gene family are functional. We have isolated a human H2B pseudogene that contains alterations in the protein-coding sequences as well as in the 3′ and 5′ flanking squences that preclude expression of a functional H2B histone protein. There are three modifications in the amino acid-coding region: a single-base deletion producing a frame shift, a single-base substitution resulting in a codon change from serine to tryptophan (an amino acid not present in histones), and the absence of a stop codon. Analysis of nucleotide sequences upstream from the AUG start signal indicates the absence of a "TATA" box and other putative consensus regulatory sequences. In the 3′ flanking region, a highly conserved block of 22 nucleotides that exhibits hyphenated dyad symmetry is displaced downstream. Within the same genomic segment, the adjacent H2A histone gene is missing 12 nucleotides, resulting in a deletion of four amino acids in a highly conserved region of the protein.**

Histone genes encode a highly conserved class of basic proteins that play a key role in the structural and possibly in the transcriptional properties of the eukaryotic genome. In human cells, the histone genes are represented as a family of moderately reiterated sequences that are clustered (1, 2) but not organized in simple tandem repeats as in *Drosophila* (3) and in sea urchin (4).

Analysis of cloned genomic human histone sequences in several laboratories (1, 2) has indicated that human histone gene clusters are polymorphic and exhibit several arrangements with respect to restriction sites and the order and representation of coding sequences, including those for H1 histones (5). Structural features shared by most histone genes include contiguous representation of histone mRNA coding sequences and a series of conserved sequences in both 3′ and 5′ flanking regions (reviewed in refs. 6 and 7).

Reasoning teleogically, the presence of 30–40 copies of histone genes provides human cells with capacity to synthesize sufficient amounts of histone protein during S phase (8-10) for packaging newly replicated DNA and also to accommodate the synthesis of specific histone variants that are not temporarily or functionally coupled with DNA replication (11). However, the potential for expression of histone polypeptides may not be predicated solely on the number of genetic sites encoding histone sequences. In this paper, we present evidence suggesting that at least one human histone gene cluster contains pseudogenes of H2B and H2A histone genes. With respect to other functional genes studied to date, the H2B gene exhibits modifications in the protein coding sequences as well as anomalies in the 3′ and 5′ flanking regions that preclude expression of a functional H2B histone protein. The adjacent H2A gene exhibits a 12-nucleotide deletion in a region encoding a highly conserved segment of the protein.

MATERIALS AND METHODS

Construction of Recombinant Plasmids. Human histone gene clusters were isolated from a λ Charon 4A library (2) and characterized; *Eco*RI restriction fragments were subcloned into the *Eco*RI site of pBR322 (12, 13).

Subcloning Strategy. The recombinant plasmid pFF435B contains the complete H2A and H2B genes of λ HHG 55 (see Fig. 1*A*) inserted between the *Eco*RI and *Hin*dIII sites of pBR322. pFF435B was digested with *Pvu* II to exclude the *Pvu* II/*Hin*dIII fragment of the insert as well as the *Hin*dIII/*Pvu*II portion of the vector containing the tetracycline-resistance region (see Fig. 1*A*) and was then recircularized. The resulting subclone, designated pFF435D, includes the complete human H2B gene and the 5′ portion of the H2A gene. The recombinant plasmids are routinely propagated in HB101.

Plasmid Isolation. Bacteria containing plasmids were grown in L broth containing ampicillin at 50 μg/ml. At cell densities equivalent to A_{590} = 0.7, chloramphenicol was added to the cultures to a concentration of 0.2 mg/ml, and incubation was continued for 13–16 hr. Isolation of form I plasmid DNA was by the cleared lysate procedure followed by CsCl/ethidium bromide gradient centrifugation as described by Clewell and Helinski (14). Ethidium bromide was removed by passage over Dowex AG1X10 (BioRad). Samples were then extensively dialyzed against 10 mM Tris·HCl, pH 8/0.1 mM EDTA at 4°C, and DNA was precipitated by addition of ethanol at −20°C in the presence of 0.1 vol of 5 M NaCl. Plasmid DNA was further purified by sieving through a BioGel A-15m column (30 × 1.5 cm) using 10 mM Tris·HCl, pH 8.0/1 mM EDTA for elution. The void volume fractions containing the plasmid DNA were pooled and ethanol precipitated.

DNA Sequence Analyses. The sequences of the DNA fragments were determined by the method of Maxam and Gilbert (15) except that the piperidine cleavage and recovery of cleaved fragments by ethanol precipitation steps followed the method of Smith and Calvo (16). Gels were dried for 20 min at 80°C on a gel dryer and air dried for 1 hr prior to autoradiography using preflashed Kodak XAR-5 film and Dupont Lightning Plus intensifying screens.

RESULTS AND DISCUSSION

Restriction maps of two human histone gene clusters (λHHG39 and λHHG55) that were isolated from a gene library cloned in λ Charon 4A (2) are shown in Fig. 1*A*. HHG55 contains genes for histones H2A, H2B, H3, and H4 while λHHG39 contains H2B and H4 histone genes (2). Our interest in the regulation of transcriptional activity of histone genes prompted us to subclone and characterize further the fragments of the human genomic clusters with respect to each individual histone gene (17). To identify the H2B his-

---

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

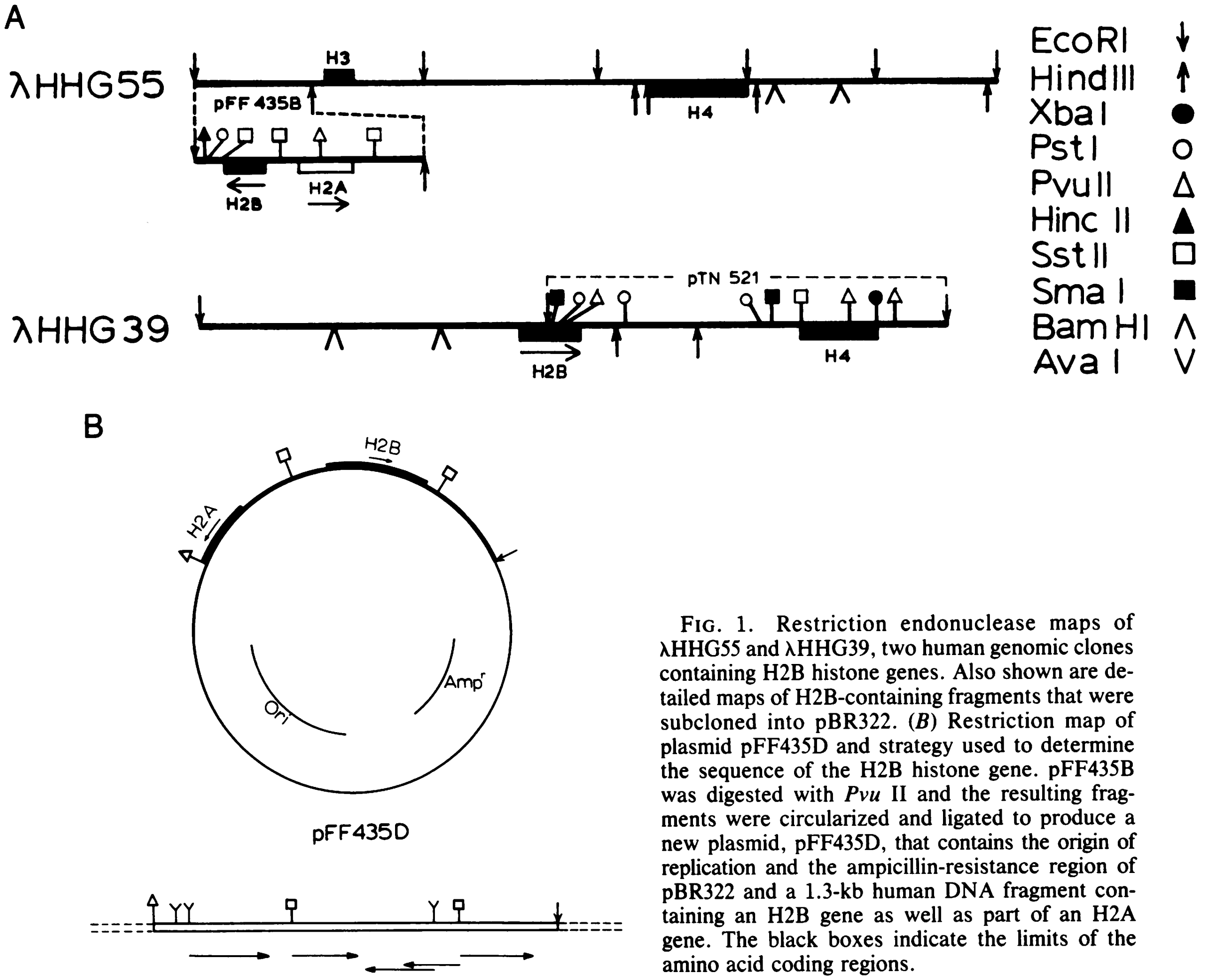

FIG. 1. Restriction endonuclease maps of λHHG55 and λHHG39, two human genomic clones containing H2B histone genes. Also shown are detailed maps of H2B-containing fragments that were subcloned into pBR322. (*B*) Restriction map of plasmid pFF435D and strategy used to determine the sequence of the H2B histone gene. pFF435B was digested with *Pvu* II and the resulting fragments were circularized and ligated to produce a new plasmid, pFF435D, that contains the origin of replication and the ampicillin-resistance region of pBR322 and a 1.3-kb human DNA fragment containing an H2B gene as well as part of an H2A gene. The black boxes indicate the limits of the amino acid coding regions.

tone subtypes encoded in λHHG55 and λHHG39 and to characterize their structural and putative regulatory components, we carried out nucleotide sequence analysis. The 2.5-kilobase *Eco*RI/*Hin*dIII fragment of λHHG55, which had been shown by hybridization-selection and *in vitro* translation to contain H2A and H2B histone genes, was subcloned into pBR322. The H2B-coding region of pFF435B was further localized to the 540-base-pair *Sst* II fragment by Southern blot hybridization of restriction endonuclease digests, using as a probe a fragment from pTN521 (Fig. 1*A*) that contains an H2B histone gene. A subclone, pFF435D, was constructed and used for DNA sequence analysis. Hybridization-selection and *in vitro* translation verified that pFF435D contains an H2B gene as well as part of an H2A gene. The strategy for DNA sequence analysis is indicated in Fig. 1*B*.

The nucleotide sequence of the human H2B histone gene from pFF435D is shown in Fig. 2, along with the sequences of H2B genes from chicken (18), yeast (19), and the sea urchin, *Psammechinus miliaris* (20). In several respects, the sequence is typical for an H2B histone gene. For example, the encoded amino acid sequence that was determined had 91% homology with that of chicken, 82% homology with sea urchin, and 71% homology with yeast. The divergence between the nucleotide sequences of the human H2B gene from pFF435D and the H2B coding regions from these species ranges from 11.0% for chicken to 26% for sea urchin and 43% for yeast. Downstream from the protein coding sequence is a highly conserved block of 22 nucleotides that exhibits hyphenated dyad symmetry capable of forming a hairpin loop structure. Although the sequence homology of this region has been retained, it is located somewhat further downstream from the end of the protein coding sequences than in H2B genes from chicken (18), *Xenopus laevis* (21) and two sea urchins (20), as well as the yeast *Saccharomyces cerevisiae* (19) (Fig. 2).

However, several features of this H2B sequence indicate that the gene is not functional. Within the 130-nucleotide sequence upstream from the AUG initiation codon, no potential "TATA" box is discernible; in other H2B histone genes that have been analyzed, the TATA region is 65–90 nucleotides 5′ to the AUG. Also, as shown in Figs. 2 and 3*a*, no stop codon was detected within 39 nucleotides of the comparable position in other H2B genes (6). Perhaps the most convincing evidence that the H2B gene from pFF435D is a pseudogene is nucleotide aberrations in the protein coding sequences. A nucleotide substitution (a C → G transversion) results in a replacement of serine 55 by tryptophan—an amino acid not present in histones. We have also observed a nucleotide deletion in the protein coding sequence at amino acid 92 (arginine) resulting in a frame shift (Fig. 3*b*). For comparison, the nucleotide sequences corresponding to amino acids 1–58 and 63–125 of the human H2B histone gene from λHHG39 are shown in Fig. 2. No anomalies have been found in this latter H2B gene.

We extended our nucleotide sequence analysis to the region of H2A gene present in pFF435D. The nucleotide sequence that codes for the amino terminus of this H2A protein and a comparison with the nucleotide sequence of chicken [λCH-01 (18)] and the protein sequences of calf, trout, and the sea urchin, *P. miliaris* (22) are shown in Fig. 4. In spite of the frequent third-base substitutions between the H2A gene in pFF435D and that in λCH-01, the polypeptide sequence is completely conserved, except for amino acids 11–14, which are absent in pFF435D. As shown in Fig. 4,

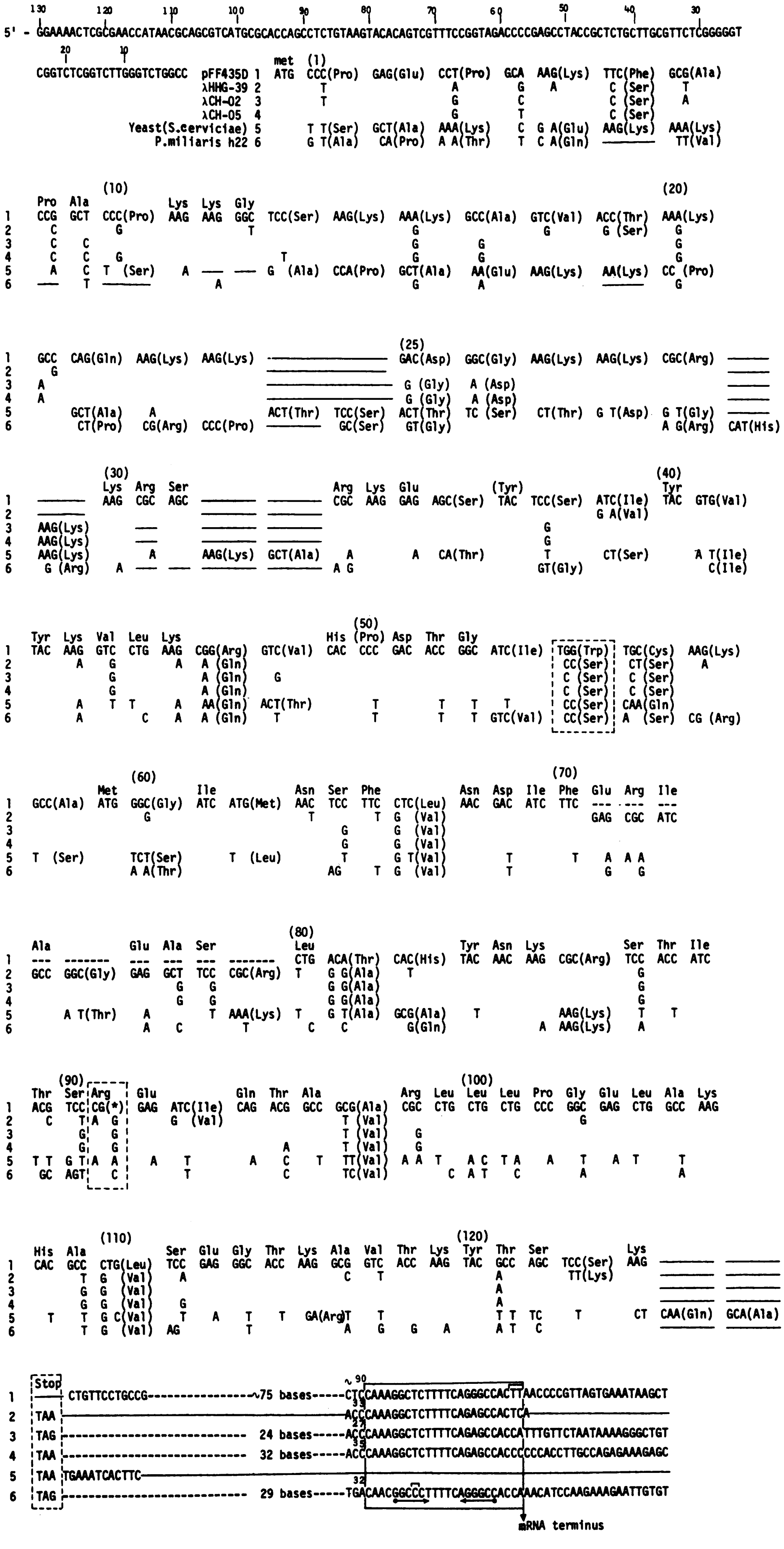

FIG. 2. Nucleotide sequence of the human H2B histone pseudogene. Shown for comparison are nucleotide sequences of another human H2B histone gene (from pTN521), two H2B genes from chicken (λCH-02 and λCH-05), an H2B type 2 gene from *S. cerevisiae*, and the sea urchin H2B gene from *P. miliaris* h22. The consensus sequence immediately preceding the mRNA terminus of the H2B genes is boxed; the limits of the hyphenated dyad symmetry are indicated by arrows.

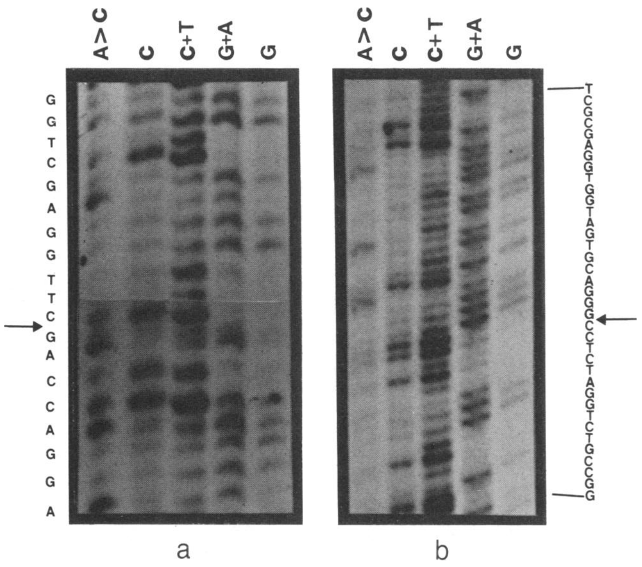

FIG. 3 (*a*) Autoradiograph of a portion of sequencing gel showing the nucleotides complementary to those coding for the final amino acids of the H2B histone protein. The last amino acid, which is lysine (triplet AAG), corresponds to the CTT immediately above the arrow. The arrow indicates the expected position of the stop codon that is absent from the sequence. (*b*) Autoradiograph showing the sequence obtained from the complementary strand corresponding to amino acids 86–96. The arrow indicates the position of a deleted cytosine that, on the complementary strand, would complete the codon for arginine (CGG).

these amino acids (Arg-Ala-Lys-Ala) are highly conserved in calf thymus, human spleen, trout, and *P. miliaris* (22). It is tempting to speculate that the absence of these four highly conserved amino acids may adversely affect the functionality of the encoded H2A protein.

Although the H2B and perhaps also the H2A histone coding sequence of λHHG55 cannot produce a functional histone protein, the functionality of the other histone genes in this cluster (H3 and H4) (Fig. 1) has not been addressed. There are two possibilities. (*i*) Actively transcribed histone genes may coexist with pseudogenes. This type of arrangement has been described for both the α- and β-globin gene clusters, where pseudogenes are interspersed with functional members of each of these gene families (23, 24). (*ii*) Alternatively, the entire histone gene cluster represented in λHHG55 may be nonfunctional. In this regard, it is noteworthy that λHHG41 (2), an independent isolate analogous to λHHG55 but lacking the H2A and H2B genes, contains H3 and H4 histone genes that, based on DNA sequence analysis (17, 25), are apparently functional. Sequence analysis of the H4 and H3 genes in λHHG55 should clearly distinguish between the above two possibilities.

The origin of the human histone pseudogene is unclear. The multiple types of defects present suggest that more than one mutational event was involved in generating this sequence, perhaps including recombination events with other histone genes. Several recent reports have suggested that pseudogenes can arise from integration of reverse transcripts of mRNAs into the genome (26-32). It is unlikely that the H2B histone pseudogene is the result of such a mechanism because it lacks the flanking direct repeats characteristic of reverse-transcribed pseudogenes. It is interesting that to date two types of arrangements of nonfunctional histone genes have been observed. In lower eukaryotes, sea urchin and *Drosophila*, solitary members of the histone gene family, designated orphons, are found at a frequency of 50 per genome (33). In contrast, the human histone pseudogenes we have identified are clustered with other human genes whose functionality has yet to be determined. Furthermore, these pseudogenes reside in a genomic segment organized in a manner similar to that of other human histone gene isolates (2).

```
           70        60        50        40        30        20        10
            |         |         |         |         |         |         |
5'-GGTAGGCAGCGGCGTTTTCGGCGCCTTTCCGATTGCCAAGCAGGAGTTTCTCTCGGTGACTACTATCGCTGTC

                                                                      (10)
              Met  Ser  Gly  Arg  Gly  Lys  Gln  Gly  Gly  Lys  Ala  ----------
pFF435D     1 ATG  TCT  GGT  CGT  GGC  AAG  CAA  GGA  GGC  AAG  GCC | ---      |
λCH-01      2        G    G    C    A         G    C    G         G | CGC(Arg) |
Calf        3                                                        | Arg      |
Trout       4                                 Thr                    | Arg      |
P. Miliaris 5                                 ---       Ala      Gly | Lys      |
    ----------------                          (20)                    ----------
   |               |  Lys  Ser  Arg  Ser  Ser  Arg  Ala  Gly  Leu  Gln  Phe
1  | ---  ---  --- |  AAG  TCG  CGC  TCG  TCC  CGC  GCT  GGC  CTT  CAG  TTC
   | Ala  Lys  Ala |
2  | GCC  AAG  GCC |                         C         C    G    G
3  | Ala  Lys  Ala |       Thr
4  | Ala  Lys  Ala |       Thr
5  | Ala  Lys  Ala |       Ser
    ----------------
                        (30)
     Pro  Val  Gly  Arg  Val  His  Arg  Leu  Leu  Arg  Lys  Gly  Asn  Tyr
1    CCG  GTA  GGG  CGA  GTG  CAT  CGC  TTG  CTG  CGC  AAA  GGC  AAC  TAC
2      C    G    C    C         C    G  C                G
3
4
5                                       Phe

   (40)                                                (50)
    Ala  Glu  Arg  Val  Gly  Ala  Gly  Ala  Pro  Val  Tyr
1   GCG  GAG  CGA  GTG  GGG  GCC  GGC  GCG  CCC  GTC  TAC
2               G         C              C    G    G
3
4
5        Asn
```

FIG. 4. Nucleotide sequence of the human H2A gene in the pFF435D subclone. The amino acids encoded by the sequence are compared with those of chicken, calf, trout, and the sea urchin, *P. miliaris*. Only the amino acid substitutions are indicated. The nucleotide sequence is also compared with that of a chicken H2A gene from λCH-01 (15).

These studies were supported by grants from the March of Dimes Birth Defects Foundation (1-813), the National Science Foundation (PCM80-18075 and PCM81-18951), and the National Institutes of Health (GM32010).

1. Heintz, N., Zernik, M. & Roeder, R. G. (1981) *Cell* **24**, 661–668.
2. Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G. & Stein, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1795–1799.
3. Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1053–1063.
4. Gross, K., Schaffner, W., Telford, J. & Birnstiel, M. (1976) *Cell* **8**, 479–484.
5. Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Wells, J. R. E., Stein, G. & Stein, J. (1984) *Science*, in press.
6. Hentschel, C. C. & Birnstiel, M. L. (1981) *Cell* **25**, 301–313.
7. Kedes, L. H. (1979) *Annu. Rev. Biochem.* **48**, 837–870.
8. Robbins, E. & Borun, T. W. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 409–416.
9. Spalding, J., Kajiwara, K. & Mueller, G. C. (1966) *Proc. Natl. Acad. Sci. USA* **56**, 1535–1542.
10. Stein, G. S. & Borun, T. W. (1972) *J. Cell Biol.* **52**, 292–307.
11. Wu, R. S. & Bonner, W. M. (1981) *Cell* **27**, 321–331.
12. Plumb, M., Stein, J. & Stein, G. (1983) *Nucleic Acids Res.* **11**, 2391–2410.
13. Sierra, F. (1983) Dissertation (Univ. of Florida, Gainesville, FL).
14. Clewell, D. & Helinski, D. R. (1970) *Biochemistry* **9**, 4428–4440.
15. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
16. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255–2274.
17. Stein, G. S., Sierra, F., Stein, J. L., Plumb, M., Marashi, F., Carozzi, N., Prokopp, K. & Baumbach, L. (1984) in *Histone Genes*, eds. Stein, G., Stein, J. & Marzluff, W. (Wiley, New York), in press.
18. Harvey, R. P., Robins, A. J. & Wells, J. R. E. (1982) *Nucleic Acids Res.* **10**, 7851–7863.
19. Wallis, J. W., Hereford, L. & Grunstein, M. (1980) *Cell* **22**, 799–805.
20. Hentschel, C. C., Irminger, J. C., Bucher, P. & Birnstiel, M. L. (1980) *Nature (London)* **285**, 147–151.
21. Moorman, A. F. M., deBoer, P. A. J., deLaff, R. T. M. & Destree, O. H. J. (1981) *FEBS Lett.* **136**, 45–54.
22. Isenberg, I. (1979) *Annu. Rev. Biochem.* **48**, 159–191.
23. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) *Cell* **19**, 959–972.
24. Lauer, J., Shen, C. K. J. & Maniatis, T. (1980) *Cell* **20**, 119–130.
25. Sierra, F., Stein, G. & Stein, J. (1983) *Nucleic Acids Res.* **11**, 7069–7086.
26. Lewin, R. (1983) *Science* **219**, 1052–1054.
27. Wilde, C. E., Crowther, C. E., Cripe, T. P., Gwo-Shu Lee, M. & Cowan, N. J. (1982) *Nature (London)* **297**, 83–84.
28. Nishioka, Y., Leder, A. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2806–2809.
29. Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D. & Leder, P. (1982) *Nature (London)* **296**, 321–325.
30. Jagadeeswaran, P. U., Forget, B. G. & Weissman, S. M. (1981) *Cell* **26**, 141–142.
31. Proudfoot, N. J. & Maniatis, T. (1980) *Cell* **21**, 537–544.
32. Lacy, E. & Maniatis, T. (1980) *Cell* **21**, 545–553.
33. Childs, G., Maxson, R., Cohn, R. H. & Kedes, L. (1981) *Cell* **23**, 651–663.